DOCUMENT RESUME

ED 218 960 FL 013 057

AUTHOR	Clark, John L. D.; Swinton, Spencer S.
TITLE	The Test of Spoken English as a Measure of Communicative Ability in English-Medium Instructional Settings. TOEFL Research Reports, 7.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RR-80-33
PUB DATE	Dec 80
NOTE	80p.
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Classroom Communication; English (Second Language); Foreign Students; Higher Education; Language Proficiency; *Language Tests; *Predictive Validity; Speech Communication; Student Evaluation of Teacher Performance; Teacher Effectiveness; *Teaching Assistants; *Test Validity; Young Adults
IDENTIFIERS	*Foreign Service Institute Interview; *Test of Spoken English

ABSTRACT

A concurrent validation analysis was conducted for the recently developed Test of Spoken English (TSE), using as an external criterion the Foreign Service Institute (FSI) direct proficiency interviewing procedure. Use-related validation data for the TSE were obtained as a predictor of the communicative effectiveness in English of non-native English-speaking teaching assistants in U.S. colleges and universities. TSE and FSI tests were administered to 134 foreign teaching assistants at nine participating institutions. The TSE subscores were somewhat more reliable than those of the FSI, and exhibited a greater degree of discriminant validity. In the use-validation phase of the study, FSI and TSE scores of 60 non-native English-speaking teaching assistants were entered as predictor variables in multiple regression analyses, using as criterion variables student ratings of the instructor's spoken language. Both TSE and FSI scores were very effective predictors of student ratings of the instructor's speaking proficiency. Somewhat lower but properly directed weightings were found for the prediction of more global aspects of the teaching performance (e.g., overall effectiveness of the instructor). It is concluded that both the TSE and FSI can predict the probable communicative facility in spoken English of non-native teaching assistants in instructional settings. (Author/SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED218960

TOEFL Research Reports

REPORT 7
DECEMBER 1980

THE TEST OF SPOKEN ENGLISH AS A MEASURE OF COMMUNICATIVE ABILITY IN ENGLISH-MEDIUM INSTRUCTIONAL SETTINGS

John L. D. Clark
Spencer S. Swinton

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Educational
Testing Service
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TOEFL
Educational Testing Service

TL013057

The Test of English as a Foreign Language (TOEFL) was developed in 1963 by a National Council on the Testing of English as a Foreign Language, which was formed through the cooperative effort of over thirty organizations, public and private, that were concerned with testing the English proficiency of non-native speakers of the language applying for admission to institutions in the United States. In 1965, Educational Testing Service (ETS) and the College Board assumed joint responsibility for the program and in 1973 a cooperative arrangement for the operation of the program was entered into by ETS, the College Board, and the Graduate Record Examinations Board. The membership of the College Board is composed of schools, colleges, school systems, and educational associations; Graduate Record Examinations Board members are associated with graduate education.

ETS administers the TOEFL program under the general direction of a Policy Council that was established by, and is affiliated with, the sponsoring organizations. Members of the Policy Council represent the College Board and the Graduate Record Examinations Board and such institutions and agencies as graduate schools of business, junior and community colleges, nonprofit educational exchange agencies, and agencies of the United States government.

The Test of Spoken English as a Measure of Communicative Ability in English-Medium Instructional Settings

John L. D. Clark

Spencer S. Swinton

Educational Testing Service
Princeton, N.J.

RR 80-33

Copyright © 1980 by Educational Testing Service.

All rights reserved. Unauthorized reproduction in whole or in part prohibited.

Abstract

The two major purposes of this study were (1) to conduct a concurrent validation analysis of the recently developed Test of Spoken English (TSE), using as an external criterion the Foreign Service Institute (FSI) direct proficiency interviewing procedure; and (2) to obtain use-related validation data for the TSE as a predictor of the "communicative effectiveness" in English of non-native English speaking teaching assistants assigned to course lecturing or other instructional roles in U.S. colleges and universities.

For the concurrent validation analysis, the TSE and FSI tests were administered to 134 foreign teaching assistants at nine participating institutions. High interrater correlations were obtained for both TSE and FSI global scores and for the available diagnostic (pronunciation, fluency, etc.) subscores on each instrument. The TSE subscores were somewhat more reliable than those of the FSI, and exhibited a greater degree of discriminant validity.

In the use-validation phase of the study, FSI and TSE scores of 60 non-native English speaking teaching assistants were entered as predictor variables in multiple regression analyses using as criterion variables student ratings of the instructor on a number of dimensions of the instructor's spoken language use in the classroom and other instructional contexts.

Both TSE and FSI scores were very effective predictors of student ratings of the instructor's speaking proficiency, with standardized beta weights of up to .63 for the TSE and .80 for the FSI. Somewhat lower but properly directed weightings were found for the prediction of more global aspects of teaching performance (e.g., overall "effectiveness" of the instructor), as measured by student responses to relevant questions and question groupings on the Student Instructional Report, a standardized instructor/course rating instrument.

Study results are considered to support the appropriateness of both the TSE and FSI testing procedures as predictors of the probable communicative facility in spoken English of non-native teaching assistants in the classroom and other typical instructional settings.

TABLE OF CONTENTS

	Page
BACKGROUND OF STUDY	1
Overview of Procedure	2
DATA GATHERING INSTRUMENTS	3
FSI Interview	3
Test of Spoken English	4
Student Instructional Report	5
Development of Language-Related Questions	6
PROCEDURES	8
Identification and Contact of Participating Institutions	8
TSE/FSI Administration	10
SIR Administration	11
DATA ANALYSIS AND RESULTS	14
TSE/FSI Intercorrelations and Scoring Reliabilities	14
TOEFL, TSE, and FSI Comparisons	18
Multiple Regression Analyses	20
FSI and TSE Scores as Predictors of Communicative Ability in English	24
Prediction of Other Instructional Factors	29
Native-English Speaking Cohort Analysis	37
Sample Expectancy Table	41
SUMMARY	43
REFERENCES	47
APPENDICES	49
A. FSI Score Levels	49
B. SIR Response Sheet	53
C. SIR Supplementary Questions	55
D. Questionnaire for Participants in the TSE Validation Study	57
E. Instructions to Students	59
F. SIR Report Form	61
G. Adjusted Squared Multiple Correlations	65

LIST OF TABLES

Table		Page
1.	Interrater Reliability--Test of Spoken English.....	15
2.	Intercorrelations Among TSE Scales--Averaged Ratings.....	15
3.	Interrater Reliability--FSI Interview.....	17
4.	Intercorrelations Among FSI Scales--Averaged Ratings.....	17
5.	TOEFL, TSE, and FSI Intercorrelations.....	19
6.	Means and Standard Deviations of Regression Analysis Variables.....	23
7.	Regression Analysis--Language-Related Variables.....	25
8.	Regression Analysis--SIR Questions.....	30
9.	Regression Analysis--SIR Factor Scores.....	33
10.	Intercorrelations of Regression Analysis Variables.....	35
11.	Correlations of School with Q39 and Q49 for Native and Non-native English Speaking TA's.....	38
12.	Correlations of Overall Effectiveness of Instructor (Q39) with Department.....	38
13.	Correlations of Overall Ability to Communicate in English (Q49) with Department.....	39
14.	Correlations of Overall Effectiveness of Instructor (Q39) with Instructional Role.....	40
15.	Correlations of Overall Ability to Communicate in English with Instructional Role.....	40
16.	Expectancy Table--Science and Math Instructors.....	42

BACKGROUND OF STUDY

The study described in the present report is an outgrowth of and closely associated with an earlier study (Clark and Swinton, 1979) which involved the initial development and experimental administration of prototype testing formats and item types for a test of English speaking proficiency that has recently been operationally introduced into the TOEFL program as the Test of Spoken English (TSE) (ETS, 1980). The procedural approach adopted in the 1979 study was to design and incorporate into each of two preliminary test forms a substantially larger number of test formats and item types than would ultimately be used in an operational instrument, and to administer these prototype tests to a representative group of non-native English speaking students. Concurrently, and to the same examinee group, was administered as an external measure the more highly face- and content-valid Foreign Service Institute (FSI) direct proficiency interview, together with, for comparison purposes, the regular Test of English as a Foreign Language (TOEFL), intended to measure listening comprehension, reading ability, and recognitional knowledge of contextually appropriate morphology and syntax. Correlational and scalar analyses were conducted on the prototype test data to identify those particular formats and item types showing the highest correlations with the FSI interview score, and relatively lower correlations with the regular TOEFL. These statistical results, together with consideration of the linguistic content and administrative requirements of the formats and question types involved, guided the preparation of a single "final version" Test of Spoken English intended for operational use.

The 1979 study thus provided initial validation data for the TSE in the form of concurrent correlation of item types appearing in the final version of the test with an external instrument (the FSI interview) considered to more directly and more self-evidently assess active speaking proficiency in real-life communicative contexts (face-to-face conversation in a variety of topical areas) than did the necessarily less direct TSE, for which the presentation of test stimuli by means of a tape recorder and printed booklet was the only operationally feasible mode of administration within the context of the TOEFL program.

Although the validity-related information provided by the initial development study was of substantial value in its own right, it was considered desirable to carry out a follow-up study of the same general type, in which the TSE--in its final operational form--would again be correlated with the FSI interview for concurrent validation purposes.

A second approach toward further validation of the TSE involved the concept of "use-validation," in which scoring results of the instrument in question are not compared to those of one or more other tests but instead to informed, independently gathered, and quantifiable judgments concerning the adequacy of the examinee's performance "in the field," that is, in the process of carrying out--in the real-life setting at issue--those particular activities for which the test is assumed to have predictive value.

In the TSE context, one of the major anticipated uses of the test was that of determining, as part of the candidate selection process, the probable communicative effectiveness, in the classroom and other typical instructional settings, of non-native English speaking applicants for U.S. college or university teaching assistantships or other instructional assignments in which English speaking ability would be a major consideration.

The primary objectives of the study reported here were thus:

(1) to carry out a concurrent validation analysis of examinee performance on the Test of Spoken English, in its present operational form, using the Foreign Service Institute interview technique as the criterion instrument; and

(2) to obtain and present a substantial amount of "use-validation" data relating the performance of non-native English speaking teaching assistants on the Test of Spoken English (as well as the FSI interview) to their actual communicative effectiveness in classroom lecturing and other settings requiring the active use of spoken English as a basic component of their instructional assignments. Both of these objectives were pursued within a single procedural design as outlined below.

Overview of Procedure

The basic approach for the study was to identify several U.S. undergraduate/graduate institutions having reasonably large numbers of non-native English speaking teaching assistants to whom would be administered the operational TSE early in the fall 1979 semester. These participants would at the same time take an FSI-type interview, to be evaluated according to the official FSI scale.

Approximately one month following the TSE and FSI administrations, the classroom students of the participating teaching assistants would be asked to evaluate, by means of specially-prepared questionnaire items, the TA's ability to communicate effectively in English, in the classroom and other instructional settings. To adjust for possible biasing effects, on the students' evaluations, of instructor characteristics not related to English language proficiency per se (for example, general course organization and planning, or the quality of the textbooks and other materials used), the students would also be asked to complete a standardized instructor/course rating form covering these and several other aspects of the teaching process. Data from these "non-language" questions would be used as statistical controls, permitting the language-related questions to be considered independently of these other variables.

As an additional precaution, native English speaking teaching "cohorts" of the non-native instructors would be identified, and the students of these cohorts also asked to rate their instructor on both communicative effectiveness in English and on the other non-linguistic

elements of the instructor/course rating form. On the assumption that native English TA's would be rated uniformly highly on these questions, and that the ratings for non-native TA's would be widely distributed along the rating scale, the cohort data would offer some operational validation of the language-rating questions. Cohort ratings on the non-language questions would help to detect (and adjust for) any "institution effect" in the rating process.

With respect to the two major research purposes of the study, concurrent administration data for the TSE and FSI would thus be obtained and the relationships of both TSE and FSI scores to the student-judged classroom performance of non-native English speaking teaching assistants determined and reported.

The three major data gathering instruments used in the study--the FSI interview, Test of Spoken English, and Student Instructional Report (including the language-use questions developed for the project)--are described in detail below, followed by a chronological description of project activities and presentation and discussion of the obtained results.

DATA GATHERING INSTRUMENTS

FSI Interview

This testing technique, usually referred to as the "FSI interview," was developed in the late 1950's by the Foreign Service Institute of the U.S. Department of State (Sollenberger, 1978; Wilds, 1975). It consists of a structured conversation of about 15-25 minutes duration between the examinee and a trained interviewer who is either a native or near-native speaker of the test language. The conversation begins at a fairly simple level and, in accordance with the overall proficiency of the examinee, becomes increasingly more sophisticated and demanding with respect to the linguistic aspects involved. The interview is continued until the interviewer is satisfied that the examinee has demonstrated the highest level of language use of which he or she is capable.

Performance on the interview is evaluated on a scale ranging from 0 to 5, with 0 representing no functional ability in the language and 5, a spoken command of the language indistinguishable in all respects from that of a native speaker. Each of the numerical score levels is accompanied by a brief verbal description of the types of real-life language-use situations in which an examinee at that level would be considered able to function in an appropriate and effective manner. For example, the verbal description of "level 1" competence is as follows:

> Able to satisfy routine travel needs and minimum courtesy requirements. Can ask and answer questions on topics very familiar to him; within the scope of his very limited language

> experience can understand simple questions and statements, allowing for slowed speech, repetition, or paraphrase; speaking vocabulary inadequate to express anything but the most elementary needs; errors in pronunciation and grammar are frequent, but can be understood by a native speaker used to dealing with foreigners attempting to speak his language. While elementary needs vary considerably from individual to individual, any person at level 1 should be able to order a simple meal, ask for shelter or lodging, ask and give simple directions, and tell time.

In addition to the verbally-defined global rating levels, a two-way grid of "Factors in Speaking Proficiency" is available to the rater. This grid provides, for each of the 5 levels, short individual characterizations of the expected scope and quality of the examinee's performance in the areas of pronunciation, grammar, vocabulary, fluency, and comprehension. These descriptions may be consulted by the rater in making the final global rating but are not an official component of the scoring process, in that individual ratings for pronunciation, grammar, etc. are not usually reported as separate scores. Appendix A shows the official verbal descriptions of each of the FSI score levels, as well as the speaking proficiency "factors" grid.

The training of FSI interview raters is an intensive process extending over several days and involves detailed study of a training manual of approximately 40 pages as well as the listening to and rating of 30 or more live and recorded interviews under the supervision of an experienced tester trainer.

Test of Spoken English

As previously described, the Test of Spoken English was developed by ETS over a three-year period as the major activity of a research project entitled "An Exploration of Speaking Proficiency Measures in the TOEFL Context." The final report of this project (Clark and Swinton, 1979) discusses the measurement rationale and procedures used in developing the TSE and describes in detail the numerous testing formats and question types investigated in the course of the study and the bases for selection of the particular subset of formats and question types included in the final version of the test.

The Test of Spoken English, in its current operational form as used in the present study, consists of seven sections, each involving a particular speaking activity on the part of the examinee. The first section is an unscored "warm-up" in which the examinee responds orally to a short series of biographical questions spoken on the test tape (name, reasons for studying English, future plans, etc.). In the second section, the examinee reads a short (about 125-word) printed passage aloud, with attention to pronunciation and overall clarity of speech. (Time is also allowed for preliminary silent reading of the passage.)

In the third section, the examinee sees a series of ten partial sentences (for example, "When the library opens...") and is asked to complete the sentence orally in a meaningful and grammatically correct way ("When the library opens, I will return the book"; "When the library opens, I will go there to study"; or other similar response).

The fourth section consists of six line drawings that "tell a continuous story" (for example, making preparations for and going out on a "night on the town"). After studying the drawings briefly, the examinee is asked to "tell the story that the pictures show," using past tense narration.

In the fifth section, the examinee looks at a single line drawing (for example, a classroom scene with one student obviously cheating on an examination) and answers a series of spoken questions about the picture ("Where is this scene taking place?" "What is the teacher doing?"), the thoughts or attitudes of the persons portrayed ("What is the teacher thinking?"), and likely future consequences of the situation ("What will probably happen to the boy?").

The sixth section consists of a series of spoken questions intended to elicit relatively free and somewhat more lengthy responses on the examinee's part. Questions requiring both straightforward descriptions of common objects (for example, "Describe a pencil in as much detail as you can") and fairly open-ended expressions of opinion (e.g., the problem of automobile pollution) are included. For the latter, solely the linguistic quality and adequacy of communication of the examinee's response are at issue in scoring, and not the factual content of the response.

In the seventh and final section, the candidate sees a printed class schedule for an imaginary course, including lecture and laboratory hours, final examination date, and other information, and is asked to describe the schedule aloud, as though informing a class.

Scoring of the TSE is carried out by trained raters, using a scoring scale based on separate evaluations of pronunciation accuracy, grammatical control, fluency, and overall comprehensibility. Pronunciation, grammar and fluency scores are reported on a scale of 0.0-3.0 and comprehensibility, on a scale of 000-300.

Student Instructional Report

The Student Instructional Report (SIR) was developed by ETS in the early 1970's as a means of obtaining, in a consistent and objective manner, student observations and opinions concerning course content and organization, teaching practices, and general instructional effectiveness of a given instructor/course situation (ETS, 1971). The development and intended utilization of the SIR are more fully described in Centra (1972). Centra (1974) and Centra and Creech (1974) report subsequent validation studies and discuss related topics.

In its present form, the SIR consists of a 39-item questionnaire, printed on two sides of an optically-scanned 8 1/2 x 11-inch sheet. The questions reflect six different course- or instructor-related factors derived from an earlier factor analytic study: (1) Course Organization and Planning (e.g., "The instructor used class time well"; "The instructor summarized or emphasized major points in lectures or discussions"); (2) Faculty/Student Interaction ("The instructor was readily available"; "The instructor made helpful comments on papers or exams"); (3) Communication ("Lectures were too repetitive of what was in the textbook(s)"; "The instructor raised challenging questions or problems for discussion"); (4) Course Difficulty and Workload (student rating of the level of difficulty of the course "for my preparation and ability"; rating of the pace at which the instructor covered the material); (5) Textbooks and Readings (rating of textbooks and supplementary readings from "excellent" to "poor"); and (6) Tests and Exams (rating of overall quality of the course examinations and the extent to which they "reflected the important aspects of the course").

Other questions not included in the above factors ask for general ratings of the value of class discussions and laboratory sessions, the "overall value of the course to me"; or touch on the student's own reasons for taking the course, anticipated final grade, and affective reactions to the course experience (extent to which "my interest in the subject area has been stimulated by this course").

The complete SIR response sheet, showing all of the regular SIR questions in the sequence and form presented to the students, is reproduced in Appendix B.

Development of Language-Related Questions

In addition to presenting the 39 regular questions, the SIR response sheet provides space for marking up to ten "supplementary questions" developed locally by the institution or individual instructors. For the present study, the "supplementary questions" section was used to present a series of questions specifically addressed to the instructor's English language proficiency and his or her ability to communicate effectively in English in classroom lecturing and other typical instructional settings.

The specific questions for this section were developed by the project staff through a series of discussion meetings, question-drafting, and joint review sessions. In preparing these questions--which were intended to serve as the basic criterion measure of instructors' "communicative effectiveness" for the study--the major developmental considerations were as follows. First, it was presumed that the student respondents would be quite unfamiliar with even the most basic linguistic terminology or shorthand descriptions of language skill that are common currency among language teachers and researchers. For example, it was felt that even the simple term "speaking proficiency" would not have an immediately meaningful

or uniform connotation for the respondents, and might be interpreted in inappropriate ways. For example, if "speaking" were interpreted in a "public address" or "formal presentation" sense, the orientation of the respondent (and the resulting answers) might be quite different from those of a respondent who interpreted "speaking proficiency" in the more general sense intended. In keeping with these considerations, it was felt that such routinely-used expressions as "speaking proficiency," "listening comprehension," "extent of vocabulary," and so forth could not be appropriately used in the student questions.

A potentially very useful approach to this problem, and one which avoided the need to phrase the questions in terms of the instructor's "listening," "speaking" etc. ability (with the attendant problems of descriptive terminology) was to cast each question in terms of the effect on the student of the instructor's language behavior. Using this "student-oriented" approach, it was possible, for example, to express a question on the instructor's speaking ability in a lecture situation in terms of the extent to which, during lectures, the instructor's English "interfered with my [i.e., the student respondent's] understanding of what was being said."

With respect to the scale on which the students would be asked to indicate their response, it was felt that greater objectivity could be obtained by asking for an appraisal of the proportion of instructional contact time during which given communication problems were evidenced, rather than making use of adjective-based descriptions of the seriousness of the problem (such as, the instructor's speech was "very difficult to understand," "somewhat difficult to understand," etc.).

The general question format finally adopted and the associated response scale are shown in the example below:

> When the instructor was lecturing to the class, his or her English interfered with my understanding of what was being said.
>
> (0) = Not applicable or don't know.
> (1) = Rarely or never.
> (2) = Occasionally.
> (3) = About half the time.
> (4) = Frequently.
> (5) = Always or almost always.

In addition to the above question on the instructor's English speaking ability in a lecture situation, similarly formatted questions were asked concerning the instructor's English use in less formal and more highly individualized contexts, including responding to students' in-class questions, communicating in one-on-one tutorial or laboratory sessions, and conversing in after-class or office-visit situations. Two additional questions were aimed at determining the general level of listening comprehension on the instructor's part: "The instructor appeared to

easily understand questions asked or statements made in class by the students"; and "When I was talking to the instructor, I had to change my own way of speaking (for example, use simpler words or talk more slowly than usual) to make sure that the instructor understood what I was saying."

Three further questions were included, which asked the students to evaluate the extent to which the instructor's "pronunciation of English," "English grammar," and "English vocabulary" interfered with comprehension. Although, as previously discussed, there was some question on the part of project staff as to whether such discriminations could reliably be made by non-linguistically trained respondents, it was decided to include them for comparative purposes. A final summary question on "the instructor's overall ability to communicate in English" completed the 10-item "Supplementary Questions" section, which is reproduced in full as Appendix C.

PROCEDURES

Identification and Contact of Participating Institutions

Initial and follow-up contacts with institutions participating in the study were made in July and early August, 1979. In identifying the schools to be approached, it was considered necessary, for reasons of administrative feasibility and cost-effective use of project staff (who would need to travel to each separate institution to conduct the FSI interviews on-site), to concentrate on those institutions having a relatively large number of potential foreign teaching assistant participants. On the basis of TOEFL staff acquaintance with institutions that were known to have fairly extensive foreign student enrollments at the graduate level (and by the same token, presumed to use a number of these students in teaching assistant positions), 15 institutions were identified and contacted to determine both their general interest in participating in the study and the approximate number of non-native English speaking teaching assistants who would be expected to be carrying out instructional assignments in the fall term. Virtually all of the contacted institutions expressed interest in the study. However, at a few institutions, anticipated difficulties in securing the cooperation of individual departments or other administrative considerations prevented their participation; at several others, the reported total number of available foreign teaching assistants was found to be fewer than the minimum of 10 established as a practical lower-bound figure by the project staff.

Formal letters of invitation were mailed to the eight institutions that were able to participate, informing them in detail of the project steps. Specifically, an identified contact person at the institution was to identify and arrange for the participation of non-native English speaking teaching assistants (most desirably in their first year of teaching in the United States) who would have a fall-term instructional assignment requiring them to "use English extensively with students,"

either teaching their own courses or leading laboratory sessions or discussion groups involving considerable interaction in English. Each of these participants would be asked to take the Test of Spoken English on a date "reasonably close to the beginning of the fall term"; at approximately the same time, a face-to-face FSI-type interview would be administered on-site by project staff. Participating teaching assistants would receive a TSE score report at no charge, as well as the FSI interview results.

As the third and final activity in the data collection process, to be carried out approximately one month after the TSE and FSI administrations, the classroom (or laboratory/discussion group) students of the participating TA's would be asked to complete the Student Instructional Report--including the ten additional questions concerning the instructor's communicative ability in English--under appropriate administration arrangements to be made at that time by the institutional contact person. For purposes of control-group comparisons, the contact person would also arrange for simultaneous SIR administration to the students of "cohort" native-English speaking teaching assistants having the same general instructional assignments as the non-native English TA's.

Complete anonymity would be maintained for all of the SIR respondents (SIR answer sheets would be identified only by a code number showing the instructor and course in question). Individually-identifiable score reports for the TSE and FSI administrations--as well as SIR results for a particular instructor/class combination--would be sent only to the individual teaching assistants involved, although the institution would receive overall score distribution data for TSE, FSI, and combined (department-level) SIR results wherever sufficient numbers of TA's were tested in a given department to provide adequate concealment of individual instructor results. Because of the large number of persons involved, no reimbursement could be offered to the participating instructors or to the classroom students. A modest honorarium was, however, provided the institutional contact for his or her activities on behalf of the project over the data gathering period.

Following receipt of the detailed informational letter but prior to the September through early-October TSE and FSI administration period, three of the eight institutions found it necessary to withdraw from the project--in two instances as a result of unanticipated difficulties in obtaining adequate departmental cooperation and, in the third instance, because of a substantial reduction in the anticipated number of TA participants.

In an effort to adjust for this situation, five additional institutions were contacted, of which four were both willing to participate and estimated an adequate number of available foreign teaching assistants. For two of these institutions, the dates of initial contact and geographic location of the institution were such as to permit their full participation in all aspects of the project, including on-site FSI testing by the project staff.

In two other instances, it was necessary to forego on-site administration of the FSI, although the other data gathering activities were possible in the regular manner.

The final total of eight participating institutions is shown below in groupings corresponding to the institutions visited by each of two project teams for the FSI administration and the chronological sequence of testing:

Iowa State University	Oklahoma State University
University of Minnesota	University of Arizona
University of Illinois	Louisiana State University
	University of Florida

University of Delaware (except FSI)
University of California (L.A.) (except FSI)

TSE-FSI Administration

At six of the seven "FSI-included" institutions, administration of both the TSE and of the individual FSI interviews took place over a two-day period between September 28 and October 6, 1979. Because of language laboratory availability, scheduling restrictions for the participating teaching assistants, and a number of other factors, it proved necessary to allow for some flexibility in test administration procedures at the individual sites. In some instances, the FSI was administered prior to the TSE and in others, the reverse sequence was followed. The time interval between the two test administrations also varied from a few minutes to overnight, again as a consequence of the scheduling limitations involved.

At a single institution, it was not possible to administer the TSE until approximately two weeks after the on-site FSI interviews, raising the theoretical possibility of slightly improved performance on the TSE (by comparison to the FSI score) attributable to additional contact with English over this period. However, since any such effect would tend to lower the observed TSE-FSI correlation (as well as the relationship with the SIR criterion) it was considered reasonable and experimentally conservative to continue to include these cases in the study.

At all seven institutions, TSE administration was carried out in a language laboratory according to the administration instructions specified in the supervisor's manual. FSI interviews were conducted on an individual basis between the teaching assistant and an ETS language staff member intensively trained in the interviewing process.[1] All interviews were

[1] At one institution, some of the interviews were conducted by two local staff members who had been trained in the interviewing technique by ETS in connection with an earlier project.

tape recorded on individual cassettes, using two small lapel microphones joined by a "y"-connector and feeding a portable cassette recorder with electronically adjusted recording level.

Basic background data on the participating teaching assistants was obtained by means of a short "Questionnaire for Participants in the TSE Validation Study" (Appendix D), which included questions on native language; number of years of English study in the native country; total number of months in the United States or other English-speaking countries; whether or not English language course(s) were being (or had been) taken in the United States; whether "any course in which the language of instruction was English" had been taught by the instructor prior to the current (fall 1979) semester; total number of years teaching "any subject in any country"; name of academic department; date and location of most recent TOEFL test; and highest academic degree received.

An additional question asked the instructor to indicate his or her departmental responsibilities for the fall 1979 term by marking all of the applicable descriptions on the following list:

I am teaching a course in (give subject) _____. The official title of the course is _____.

I lead a discussion section after the professor lectures.

I assist in laboratory sessions (help the students with equipment, answer questions, and so forth).

I discuss their work with individual students (tutorial sessions).

I grade student papers and/or examinations for a professor or another instructor.

I assist a professor in doing research.

Other responsibility (please describe) _____.

SIR Administration

Approximately three weeks after the FSI and TSE administrations on-site (for two institutions, TSE only), project staff forwarded SIR answer sheets and associated materials to the contact person at the nine participating institutions. To facilitate materials handling on-site and to insure that all institutions would carry out the SIR administration in the same manner, as much as possible of the necessary materials packaging and identification was done ahead of time by the project staff. Specifically, a manila envelope containing 35 blank SIR answer sheets was prepared for each non-native English speaking instructor at the institution who had (1) previously taken the TSE (and, in most cases, an FSI interview), and (2) indicated on the "Questionnaire for Participants"

form (Appendix D) that he or she had one or more of the following responsibilities: teaching a course; leading a discussion section; or assisting in laboratory sessions.

Respondents who indicated only that they discussed work with individual students, graded student papers, or assisted a professor in research were not felt to have sufficient communicative contact with students to warrant including them in the SIR analysis portion of the study (although their FSI and TSE data were included in the scoring reliability determinations for both tests).

If for any reason an appropriate native-English speaking cohort could not be obtained, SIR administration was nonetheless to be carried out in the non-native English instructor's class. A four-digit identification number identical to the corresponding non-native English speaking instructor's number except for one ("English" vs. "non-English") digit was provided on the cohort SIR envelopes.

Included in each non-native English and cohort envelope of SIR answer sheets were 35 copies of the "Supplementary Questions" sheet giving the ten language-specific questions developed for the study (Appendix C), as well as a sheet of administration instructions which provided information on distributing and collecting the SIR's, together with background and orientation information to be read aloud to the students (Appendix E). In these instructions, the students were assured that their responses on the SIR would be completely anonymous, and that "your own answers will not be made available to your instructor or to other persons at the institution," nor have "[any] effect whatsoever on your course grade or any other aspects of your course work." It was further indicated that "the [SIR] answers will not be used to evaluate your instructor, and information identifying your instructor will not be released. Therefore, a frank report will benefit the overall teaching at your institution but can neither benefit nor harm individual instructors."

The students were asked to "answer all questions [including the Supplementary Questions] in terms of your instructor's teaching, lab sessions, or other instructional contacts up to this point in the course. The instructor on which your answers should be based is [name supplied] and the course is [course title supplied]."

On the outside of the individual instructor envelopes, a label was affixed showing the name of the instructor and, if the latter information had previously been provided by the instructor on the "Questionnaire for Participants," the course taught. The contact person was asked in each instance to verify the accuracy of the "course taught" information or, in the case of instructors who indicated only that they led discussion or laboratory sessions, to supply the relevant course identification.

A four-digit number was also shown on the envelope label, uniquely identifying the particular institution and instructor. This number was to be used by the individual students to identify the SIR forms which they filled out concerning that instructor.

In addition to the SIR envelopes prepared for each of the non-native English speaking instructors, a similar envelope was prepared for a "native English speaking cohort" of that instructor, who was to be identified by the local contact person, with the assistance of the department chairman or other course coordinator as necessary. Desired criteria for the native English cohorts were described in the written instructions to the institutional contact persons as follows (in order of decreasing importance):

1) In same department.

2) Teaching same course. (If the non-native English TA is not lecturing in the course, it would be desirable to identify a native English TA whose only responsibility for that course is lab work, discussion session, or other activity indicated by the non-native TA. In the absence of such a close pairing, however, it would be acceptable to identify, for example, a "lecturing" native English TA to pair up with a "lab session" non-native English TA, provided that the course in question is the same.)

3) Same amount of teaching experience at the institution. To the extent possible, "first-year" non-native English TA's should be paired with "first-year" native TA's and similarly for second-year or even more experienced TA's.

Following administration of the SIR's to both non-native English speaking instructors and native English cohorts, the SIR answer sheets were returned to ETS in the individual identifying envelopes and scored by the standard optical scanning procedure used in the SIR program. This provided, for individual instructor/course combinations: percentage distributions of responses and mean scores for each of the SIR questions, including the language-related supplementary questions; and factor scores for six factor-analytically determined groupings of SIR questions: Course Organization and Planning, Faculty/Student Interaction, Communication, Course Difficulty and Workload, Textbooks and Readings, and Tests and Exams.

The scoring program also provided similar summary (across-instructor) data based on academic department groupings. These summaries were later forwarded to the participating institutions for all situations in which two or more instructors were included in the grouping. (SIR summary reports for "groups" of only a single instructor were not provided, since this would have permitted associating the SIR results with individually identifiable instructors).

Appendix F reproduces the SIR report form and shows the particular questions comprising each of the six factors.

DATA ANALYSIS AND RESULTS

TSE/FSI Intercorrelations and Scoring Reliabilities

As previously discussed, one of the two major purposes of the study was to determine the nature and extent of statistical correspondence between the more highly face- and content-valid Foreign Service Institute direct proficiency interview and the "semi-direct" Test of Spoken English which, for practical administrative reasons, makes use of booklet-and-tape recorded stimuli (and recorded responses) rather than "live" face-to-face conversation. A high degree of intercorrelation between the FSI and the TSE would support use of the latter instrument as a reasonable and effective alternative to the FSI interview for situations in which face-to-face testing would not be operationally feasible.

Since the intercorrelation of any two measures is affected by the reliability of the individual measures, it was considered desirable to examine first the reliability of the scoring procedures for both TSE and FSI, as shown in Tables 1 and 3, respectively.

Interrater reliability figures shown for the TSE (Table 1) are based on the independent scoring of a given test tape by each of two raters. The underlined correlations on the main diagonal provide estimates of the interrater reliability of the Comprehensibility score for the TSE and of the Pronunciation, Grammar, and Fluency scores. Some evidence for discriminant and convergent validity of the scoring scales on which these results are based is seen in that each score correlates more highly with "itself" (as rated by a second rater) than it does with any of the other scores.

All of the TSE reliabilities have high absolute values, ranging from .77 to .85. This is probably attributable in large part to the discrete nature of the TSE scoring procedure, in which separate scoring judgments are made for each of the item type sections in the test and, where applicable, for individual items within sections.

It is important to note that in operational scoring of the TSE (i.e., for test scores reported to candidates) all test tapes are routinely evaluated by two separate raters and the reported score is based on an average of the two ratings. Thus, the reliabilities shown in Table 1 (which represent the estimated reliability of a single rater) should be interpreted as lower bound figures, giving a conservative estimate of the reliability of the operational TSE scoring process.

To investigate further the intercorrelations among the four TSE scores (Comprehensibility, Pronunciation, Grammar, and Fluency), the two individual ratings of each TSE tape were averaged and the correlations of these averages obtained, as shown in Table 2. These figures indicate that the general Comprehensibility rating is more closely related to Pronunciation and Fluency (r=.92 and .91, respectively) than to Grammar (r=.84), an outcome that is in keeping with the analytic process and item

Table 1

Interrater Reliability--Test of Spoken English

(N = 134)

		Rater 1			
		Comprehensibility	Pronunciation	Grammar	Fluency
Rater 2	Comprehensibility	.79	.76	.73	.76
	Pronunciation	.74	.77	.69	.73
	Grammar	.74	.69	.85	.72
	Fluency	.74	.71	.73	.79

Table 2

Intercorrelations Among TSE Scales--Averaged Ratings

(N = 134)

	Pronunciation	Grammar	Fluency
Comprehensibility	.93	.84	.91
Pronunciation		.79	.88
Grammar			.82

type selection procedures used in developing the TSE. (See Clark and Swinton, 1979 for detailed description.)

For purposes of the study, the FSI raters were asked to provide, in addition to the global rating, subscores for Pronunciation, Grammar, Vocabulary, Fluency, and Comprehension, as based on the descriptions in the "Factors in Speaking Proficiency" grid (Appendix A). For the global rating, "plus" values, where appropriate, were numerically coded as .7 (for example, 1+ = 1.7). For the grid-based subscores, the raters were permitted to rate "between" adjacent grid descriptions when they felt that the examinee's performance was at an intermediate level between the two descriptions; these intermediate ratings were also coded as .7. (For example, with reference to the grid descriptions, a control of vocabulary that the rater considered to lie between "adequate for simple social conversation and routine job needs" and "adequate for participation in all general conversation and for professional discussions in a special field" would be represented as 2.7 for the Vocabulary subscore.)

Scoring reliabilities obtained for the FSI interview are shown in Table 3, both for the total ("global") score and for separate ratings of the five component factors. Again, reliabilities are shown on the main diagonal, but unlike the TSE subscores, convergent-divergent validity assumptions are not upheld for all of the subscore comparisons. Although the appropriate convergent-divergent pattern is shown for pronunciation and grammar, higher correlations with one or more of the other subscales than with the subscale itself are found for vocabulary, fluency, and comprehension, suggesting some lack of conceptual and operational independence among these three factors, at least insofar as they are reflected in the FSI scoring process for the "grid" descriptions.

The total FSI score and the TSE Comprehensibility score are equal in reliability (r=.79), but the FSI subscales for pronunciation, grammar, and fluency are less reliable than the counterpart TSE scales, especially for pronunciation (.59 vs. .77). The apparent superiority of the TSE in rating pronunciation is probably attributable to two factors. First, the relatively tighter control of the examinee's responses in the TSE (including the reading of a printed paragraph aloud) would provide a much more uniform basis for judgments of pronunciation accuracy than would the more "free-form" FSI situation. Second, the relatively slight weight that is given to pronunciation accuracy in the FSI scoring system (once a level of sheer comprehensibility has been reached) could make the FSI raters somewhat less sensitive to this particular aspect of examinee performance.

In the FSI interview, the relatively greater freedom that the examinee has to "pick and choose" the lexical items which he or she uses in the conversation may also reduce across-examinee variance (and hence, scoring reliability) for vocabulary by comparison to the TSE situation, in which all examinees are forced to deal with similar, pre-specified topical areas.

Table 3

Interrater Reliability--FSI Interview

(N = 94)

		Rater 1					
		Global Rating	Pron.	Gram.	Voc.	Fluency	Comp.
Rater 2	Global	.79	.63	.82	.70	.72	.82
	Pronunciation	.58	.59	.57	.46	.52	.54
	Grammar	.77	.58	.80	.67	.66	.79
	Vocabulary	.72	.52	.76	.64	.66	.75
	Fluency	.70	.55	.71	.62	.65	.75
	Comprehension	.70	.52	.73	.63	.67	.76

Table 4

Intercorrelations Among FSI Scales--Averaged Ratings

(N = 94)

	Pron.	Gram.	Voc.	Fluency	Comp.
Global Rating	.74	.96	.93	.92	.90
Pronunciation		.74	.66	.69	.67
Grammar			.91	.87	.89
Vocabulary				.92	.90
Fluency					.92

As shown in Table 4, grammar is the FSI subscale most highly associated with the overall FSI score--the reverse of the case for the TSE Grammar/Comprehensibility relationship. This is consistent both with the original development process for the TSE--in which items were selected to maximize loadings on pronunciation and fluency--and with the considerable weight that is given to morphological and syntactical accuracy in the verbal descriptions (and rating process) for the FSI interview.

The contrast of correlations of component subscores and overall score for the TSE and FSI interview indicates rather clearly that the two tests emphasize somewhat different aspects of spoken language production that should be taken into account by the potential user, both in selecting an appropriate instrument for a given application and in analyzing the testing results. However, notwithstanding the particular differences cited above, the quite high total score correlation between the FSI and TSE obtained in the study (.80) suggests that while the two instruments are not identical in the aspects of language they measure, the degree of overlap is sufficient to warrant consideration of the TSE as a reasonable alternative to the FSI interview when it is not possible to carry out face-to-face testing. The TSE could also, of course, be considered for primary use in its own right, especially when accuracy of pronunciation (as well as overall fluency and comprehensibility) is an important component of the information that is desired from the testing.

TOEFL, TSE, and FSI Comparisons

Because of the rather substantial demands that would be made on the participating non-native English teaching assistants' good will and time in taking the Test of Spoken English, completing the "Questionnaire for Participants," and taking an individual FSI interview, it was not considered feasible to add the further requirement of a TOEFL administration on-site. As a less rigorous approach to obtaining comparative information for the TOEFL, but one that was considered able to provide at least some informational value, it was decided to retrieve (with the necessary permissions) the prior TOEFL score records of the participating teaching assistants and to incorporate in a TSE-FSI-TOEFL comparison any TOEFL scores that could be considered recent enough to represent the general ability level of the teaching assistant as of the fall 1979 period of the study. It was decided that scores up to one year old (i.e., from tests administered in September 1979 or later) could be reasonably used for this purpose.

Examination of the available score records showed that only 34 of the total of 137 participating teaching assistants had TOEFL scores meeting this criterion--an insufficient number to permit confident and detailed analysis. Nonetheless, within the acknowledged limitations of the sample size and the corresponding cautions on interpretation which this imposes, it was considered reasonable to obtain an intercorrelation matrix of TSE, FSI, and TOEFL scores for this group as a possible indication of general trends which could be investigated further in connection with other

TSE-related studies. These correlations are shown in Table 5, based on 31 cases having all of the necessary score data (TOEFL score less than one year old, TSE score, and FSI interview rating).

Table 5
TOEFL, TSE, and FSI Intercorrelations
(N = 31)

	TOEFL I	TOEFL II	TOEFL III	TOEFL TOTAL	TSE PRON.	TSE GRAM.	TSE FLU.	TSE COMP.	FSI TOTAL
TOEFL I	1.00	.77	.67	.92	.68	.76	.65	.69	.71
TOEFL II	.77	1.00	.64	.91	.42	.54	.52	.46	.57
TOEFL III	.67	.64	1.00	.85	.38	.56	.43	.36	.62
TOEFL TOTAL	.92	.91	.85	1.00	.56	.70	.60	.57	.71
TSE PRON.	.68	.42	.38	.56	1.00	.86	.92	.95	.77
TSE GRAM.	.76	.54	.56	.70	.86	1.00	.89	.88	.73
TSE FLU.	.65	.52	.43	.60	.92	.89	1.00	.93	.76
TSE COMP.	.69	.46	.36	.57	.95	.88	.93	1.00	.76
FSI TOTAL	.71	.57	.62	.71	.77	.73	.76	.76	1.00

Considering FSI Total as an external criterion of general speaking proficiency, it may be noted that the TSE Comprehensibility score and the three Pronunciation, Grammar, and Fluency scores correlate more highly with the FSI than do the TOEFL total score or any of the three TOEFL subscores of Listening Comprehension, Vocabulary and Structure, and Reading Comprehension. Of the three TOEFL subscores, and as would be expected, the Listening Comprehension score is the most highly correlated with FSI.

Among the three TSE subscores, TSE Grammar is more closely associated with TOEFL total (and with each of the TOEFL subscores) than are both TSE Pronunciation and TSE Fluency, suggesting that the latter two scores of the TSE are tapping somewhat different aspects of the examinee's performance

than are either the TOEFL test as a whole or the Grammar score of the TSE. The two lowest correlations in the matrix are TSE Pronunciation vs. TOEFL Reading Comprehension (Section III); and TSE Comprehensibility vs. TOEFL Reading Comprehension, which is again quite in keeping with the content and intended functioning of these tests/test sections.

Multiple Regression Analyses

The second set of analyses addressed the question of the extent to which instructor scores on the TSE and FSI--each considered separately--would contribute to the prediction of student-judged communicative ability in English on the instructor's part, over and above the predictive value that might be provided by other available measures such as the number of English language courses previously taken by the instructor or length of residence in the United States. These analyses also simultaneously probed the relationships between TSE and FSI scores and other more general ratings of instructional effectiveness (e.g., "faculty-student interaction") which would be expected to involve a language proficiency component to some extent but reflect the contribution of a number of other variables as well.

To carry out these analyses, multiple regressions were calculated for the entire group of teaching assistants (across institutions) for whom both FSI and TSE scores were available (N=60). Regressions were obtained with the FSI score and other relevant independent variables as predictors of average class ratings on each of the six SIR factors (FS1 through FS6), those regular SIR questions not part of a factor (identified as Q[no.]), and the ten specific questions on communicative effectiveness in English developed for the study (Q40-Q49). Identical regressions, but with the TSE score used in place of the FSI score as a predictor, were also run for the same group.

In both regressions, a number of other independent (predictor) variables in addition to the FSI or TSE scores were entered, according to the amount of variance explained, from three data sources. First, instructor-reported personal background variables were drawn from the teaching assistants' responses to the "Questionnaire for Participants" (Appendix D), including: number of months the teaching assistant had spent in the U.S. or other English-speaking country (US MONTHS); whether or not any English language course(s) were currently being taken or had been taken in the U.S. (US ENG NOW); if "yes" to the preceding, the total number of weeks of such language study (US ENG WKS); years of English language study in the native country (ENG STUDY); whether or not the instructor had previously taught a course using English (PRIOR ENG); highest academic degree received (HIGHEST D); number of years of teaching experience in any language (YRS TEACH); and current teaching role (ROLE 1), coded dichotomously as teaching own course vs. serving as discussion leader or laboratory assistant.

Second were added categorical variables representing the institution at which the instructor was serving (coded SCHOOL 1 to SCHOOL 9) and academic department involved. For the latter, to facilitate analysis and provide sufficiently large N's in each category, academic department memberships were combined into the three general categories of Mathematics and Science (MATH/SCI), Engineering (ENGINEER), Foreign Language (FOR LANG), and other (OTHER DEP).

Third, the SIR questionnaire yielded predictors identified in previous research (Centra and Creech, 1974) as variables related to student ratings but beyond the control of the teacher, including appropriateness of class size for the teaching method used (Q25; here and following, see correspondingly numbered questions on SIR form, Appendix B, for exact wording of questions); whether or not the course is in the student's major field (dichotomous coding derived from Q26); whether or not the course is required (Q27); the grade expected by the student in that course (Q28); the student's current grade-point average (Q29); student's year in school (freshman through graduate student: Q30); and sex of student (Q31). As entry data into the regressions, each of the independent variables was averaged over all students rating a given instructor.

This last set of independent variables, because it is taken from the same student response form (and same respondents) from which the criterion ratings of instructor competence were obtained, may be expected to have a somewhat inflated estimated predictive strength. In particular, students' judgments of the appropriateness of class size and of their expected grade may be influenced by, as well as influence, their judgments of teacher effectiveness. To the extent that such counfounding (and possible halo) effects took place, the regression analyses would represent an underestimate of the residual predictive validity of the FSI and TSE.

Table 6 shows the means and standard deviations, and intercorrelations obtained for all of the variables included in the regression analyses. For ease of reference, the variable codes used in the table are given below with a brief description of each variable:

FSI TOTAL	FSI Global Rating
TSE COMP	TSE Comprehensibility Score

TA Background Questionnaire Data

US MONTHS	Months in U.S. or other English-speaking Country
US ENG WKS	Weeks of English Study in U.S.
US ENG NOW	Currently Taking English Course
ENG STUDY	Years of English Study in Native Country
PRIOR ENG	Prior Course Taught Using English
HIGHEST D	Highest Academic Degree Received
YRS TEACH	Years Teaching in Any Language
ROLE 1	Teaching vs. Discussion or Lab Assistant
ENGINEER	Engineering Department
MATH/SCI	Mathematics or Science Department
FOR LANG	Foreign Language Department
OTHER DEP	Other Departments

Student SIR Responses

Q25	Class Size is Appropriate
Q26	Course is in Major Field
Q27	Course is Required
Q28	Expected Grade in Course
Q29	Current Grade-Point Average
Q30	Class Level (Freshman-Graduate)
Q31	Sex
Q35	Overall Quality of Lectures
Q36	Overall Quality of Class Discussions
Q37	Overall Quality of Laboratories
Q38	Overall Value of Course
Q39	Overall Effectiveness of Instructor

Student SIR Responses: Language-Specific Questions

Q40	Instructor's English Interfered with Understanding Lectures
Q41	Instructor Understood Student Questions and Statements
Q42	Instructor's English Made Answers to Questions Unclear
Q43	Easy to Understand Instructor in One-on-one In-class Situations
Q44	Trouble Understanding Instructor in Private Conversations
Q45	Had to Change Own Speech so Instructor Would Understand
Q46	Instructor's Pronuncation Interfered with Understanding
Q47	Instructor's Grammar Interfered
Q48	Instructor's Vocabulary Interfered
Q49	Instructor's Overall Ability to Communicate Interfered

SIR Factor Scores

FS1	Course Organization and Planning
FS2	Faculty-Student Interaction
FS3	Communication
FS4	Course Difficulty and Workload
FS5	Textbooks and Reading Assignments
FS6	Tests and Exams

For the regression analyses which follow, all of the figures shown are based on a total of 60 teaching assistants provided by Schools 1-7. (Schools 8 and 9 did not administer the FSI and are thus excluded from this portion of the analysis.) Within this sample, 32% of the teaching assistants were from School 3, ranging down to 5% from School 1. Fifty-eight percent of the sample was from Mathematics or Science Departments, 12% from Engineering, 10% from Foreign Language Departments, and 20% from other departments. Eighty percent reported that they were "teaching a course" in the fall semester and 20 percent indicated that they were not teaching their own course but served as discussion leaders or laboratory assistants.

Table 6

Means and Standard Deviations of Regression Analysis Variables

Variable	Mean	Standard Deviation
FS1	9.01	1.01
FS2	9.02	0.95
FS3	9.12	0.97
FS4	9.21	0.95
FS5	8.95	1.12
FS6	9.07	1.26
Q35	3.13	0.72
Q36	2.92	0.59
Q37	2.92	0.68
Q38	3.19	0.53
Q39	2.86	0.61
Q40	2.32	0.70
Q41	3.72	0.57
Q42	2.23	0.67
Q43	3.81	0.53
Q44	1.86	0.56
Q45	1.84	0.66
Q46	2.22	0.61
Q47	1.88	0.51
Q48	1.84	0.52
Q49	2.10	0.62
Q25	0.67	0.22
Q26	3.02	0.60
Q27	0.74	0.26
Q28	3.07	0.34
Q29	5.66	0.55
Q30	2.30	0.84
Q31	1.40	0.23

Variable	Mean	Standard Deviation
MATHSCI	0.58	0.50
ENGINEER	0.12	0.32
FORLANG	0.10	0.30
OTHERDEP	0.20	0.40
SCHOOL1	0.05	0.22
SCHOOL2	0.12	0.32
SCHOOL3	0.32	0.47
SCHOOL4	0.12	0.32
SCHOOL5	0.17	0.38
SCHOOL6	0.17	0.38
SCHOOL 7	0.07	0.25
TSECOMP	225.33	51.60
FSITOTAL	3.29	0.83
ENGSTUL?	9.55	4.67
USMONTHS	24.55	25.07
USENGNOW	0.47	0.50
USENGWKS	6.68	11.16
PRIORENG	0.45	0.50
YRSTEACH	2.37	1.18
ROLE1	0.80	0.40
HIGHESTD	1.48	0.50

The group averaged two years of residence in the U.S. and had over 9.5 years of English study in their home countries. Forty-five percent claimed prior teaching experience using English, and the average number of years of teaching experience was 2.37. Although this group in general thus reported themselves to be quite experienced, 47 percent were currently studying English, and the mean weeks of English study in the U.S. was only 6.7, although this last distribution was bimodal. The mean FSI interview score was slightly beyond the "3" level (3.29), with a standard deviation of .83 (range 1.0 to 4.7). The mean TSE Comprehensibility score was 225.33, with a standard deviation of 51.60 and range of 93 to 300.

FSI and TSE Scores as Predictors of Communicative Ability in English

On the basis of the relatively high mean FSI and TSE scores shown by the teaching assistants in the study, a "ceiling effect" might be anticipated, in which the average level of English proficiency would be sufficiently high that differences in language competence would be less important than command of subject matter or teaching methodology in predicting student ratings of instructor performance. Inspection of the data, however, shows that both the FSI and TSE scores constitute remarkably strong predictors of student-rated communicative facility in English, even for this generally highly competent instructor group.

Table 7 shows the regression analyses for the ten "language-specific" SIR questions developed for the study, as predicted by both FSI and TSE scores. For each question, the complete vectors of standardized beta-weights (B) and F-values for all variables entered are given in the table, and should be consulted for detailed analysis. The most salient results for each of the "language-specific" questions are discussed under the individual question headings below.

Use of English in lecture situations (Q40). Student ratings for the question "When the instructor was lecturing to the class, his or her English interfered with my understanding of what was being said" were predicted with an R^2 of .66 for the FSI analysis and .60 for the TSE analysis. In each case, by far the strongest individual predictor was the FSI (B = -.63) or TSE score (B = -.52). Years of English study in the native country (ENG STUDY) was a much less effective predictor of instructors' English use in lecture situations, with beta weights of only .26 and .12 for the FSI and TSE analyses.

Comprehension of students' in-class questions and statements (Q41). This question concerned the instructor's ability to "easily understand questions asked or statements made in class by the students" as a measure of listening comprehension in the classroom setting. This variable was very well predicted in both the FSI (R^2 =.61) and TSE (R^2 =.54) analyses. Again, for both analyses, the direct tests of speaking ability were appreciably better predictors of this aspect of English language facility than were any of the instructor background or other variables analyzed (beta-weights of .69 and .46 for FSI and TSE, respectively).

Table 7

Regression Analysis--Language-Related Variables

FSI Analysis			Q40 Instructor's English Interfered with Understanding Lectures		Q41 Instructor Understood Student Questions and Statements		Q42 Instructor's English Made Answers to Questions Unclear		Q43 Easy to Understand Instructor in One-on-one In-class Situations		Q44 Trouble Understanding Instructor in Private Conversations		Q45 Had to Change Own Speech so Instructor Would Understand		Q46 Instructor's pronunciation Interfered with Understanding		Q47 Instructor's Grammar Interfered		Q48 Instructor's Vocabulary Interfered		Q49 Instructor's Overall Ability to Communicate Interfered	
			R^2 = .66		R^2 = .61		R^2 = .61		R^2 = .46		R^2 = .66		R^2 = .73		R^2 = .72		R^2 = .62		R^2 = .71		R^2 = .72	
			Beta	F	Beta	F	Beta	F	Beta	F	Beta	F	Beta	F	Beta	F	Beta	F	Beta	F	Beta	F	Beta	F
FSI Total	FSI Total		-.62	43.60	.69	43.44	-.57	30.20	.48	15.66	-.54	27.98	-.53	27.61	-.70	50.77	-.76	48.68	-.80	51.80	-.69	55.00
US MONTHS	Months in U.S. or English-speaking Country		-.23	5.90			-.21	3.91	.16	1.75			-.23	3.51	-.14	2.30			-.11	1.28	-.12	1.21
US ENG WKS	Weeks of English Study in U.S.		-.13	1.85	.21	2.65	-.15	2.11					-.12	1.55	-.22	6.93	-.20	4.45	-.20	5.03	-.26	8.60
US ENG NOW	Currently Taking English Course				.11	0.74																
ENG STUDY	Years of English Study in Native Country		.26	7.18	-.18	3.05	.18	3.09	-.28	5.64			.22	4.98	.14	2.57	.14	2.11	.17	3.54	.12	1.73
PRIOR ENG	Prior Course Taught Using English										-.24	4.84	-.12	1.11								
HIGHEST D	Highest Academic Degree Received				.13	1.94											-.11	1.22	-.12	1.74		
YRS TEACH	Years Teaching in Any Language										.22	4.82	.19	2.95			-.13	1.98				
ROLE 1	Teaching vs. Discussion or Lab Assistant								.09	0.68	-.24	6.92										
Dept.	ENGINEER	Engineering Department			-.25	3.74					.29	6.03	.32	5.92								
	MATH SCI	Mathematics or Science Department																	-.21	3.51		
	FOR LANG	Foreign Language Department	-.30	7.81			-.32	7.52	.26	3.62					-.18	3.98					-.11	1.41
	OTHER DEP	Other Departments											.26	2.77	.15	1.81	.28	4.11				
	School 1		.11	1.55	-.20	3.97	.14	1.97													.25	4.73
	School 2		-.28	9.94	.24	5.91	-.27	7.77	.20	3.06			-.11	0.69	-.36	16.96	-.17	2.67	-.35	13.06	-.05	0.12
	School 3								.14	1.52			-.06	0.12	-.20	4.99			-.26	3.84	.29	2.51
	School 4						-.13	2.00					-.17	1.54	-.29	7.36	-.15	1.71	-.28	7.04	.03	0.07
	School 5										.14	1.80	.29	4.19			.15	2.29			.24	3.08
	School 6		.25	7.87			.25	6.70			.29	7.04	.47	9.05			.23	5.18			.21	2.03
	School 7														-.12	1.96			-.23	6.35		
	School 8																					
	School 9																					
Average	SIRNNQ 25	Class Size is Appropriate											.10	0.73								
Responses	SIRNNQ 26	Major vs. Minor vs Elective Field									.18	1.80										
for	SIRNNQ 27	Course is Required	-.13	1.60	.10	1.13	-.12	1.11	.16	1.66	-.15	1.57	-.16	2.49								
Students	SIRNNQ 28	Expected Grade											-.17	1.59							.18	2.80
Rating	SIRNNQ 29	Grade Point Average									-.13	1.25	.21	2.01					-.02	0.02	-.25	3.58
This	SIRNNQ 30	Class Level (Freshman-Graduate)			.17	.23					-.21	2.76										
TA	SIRNNQ 31	Sex			-.26	5.72			.17	2.13	.15	2.05	.27	6.56					.15	3.05	.09	1.19

Table 7 (cont.)

TSE Analysis

		Q40		Q41		Q42		Q43		Q44		Q45		Q46		Q47		Q48		Q49	
		Instructor's English Interfered with Understanding Lectures		Instructor Understood Student Questions and Statements		Instructor's English Made Answers to Questions Unclear		Easy to Understand Instructor in One-on-one In-class Situations		Trouble Understanding Instructor in Private Conversations		Had to Change Own Speech so Instructor Would Understand		Instructor's pronunciation Interfered with Understanding		Instructor's Grammar Interfered		Instructor's Vocabulary Interfered		Instructor's Overall Ability to Communicate Interfered	
		R^2 = .60		R^2 = .54		R^2 = .60		R^2 = .46		R^2 = .59		R^2 = .61		R^2 = .67		R^2 = .48		R^2 = .56		R^2 = .61	
		Beta	F	Beta	F	Beta	F	Beta	F	Beta	F	Beta	F	Beta	F	Beta	F	Beta	F	Beta	F
TSE COMP	TSE Comprehensibility	-.52	26.55	.46	15.54	-.45	17.69	.50	17.61	-.37	11.31	-.37	13.24	-.57	35.04	-.44	13.48	-.56	29.91	-.63	39.71
US MONTHS	Months in U.S. or English-speaking Country	-.25	5.27			-.19	2.98	.19	2.48			-.15	1.94	-.14	1.97			-.14	1.92	-.19	2.54
US ENG WKS	Weeks of English Study in U.S.			.19	2.94									-.12	1.69					-.11	1.27
US ENG NOW	Currently Taking English Course																				
ENG STUDY	Years of English Study in Native Country	.12	1.42					-.21	3.55												
PRIOR ENG	Prior Course Taught Using English			.23	3.38					-.33	7.99	-.27	6.26								
HIGHEST D	Highest Academic Degree Received									-.16	2.14							-.11	1.13		
YRS TEACH	Years Teaching in Any Language			-.20	3.14					.33	9.18	.31	9.38	.12	1.91			.16	2.48		
ROLE 1	Teaching vs. Discussion or Lab Assessment									-.14	1.81										
D ENGINEER	Engineering Department			-.16	1.30					.14	1.24										
e MATH SCI	Mathematics or Science Department															.25	4.09				
p FOR LANG	Foreign Language Department	-.41	12.70			-.35	11.66	.34	6.88	-.14	1.17			-.38	12.20					-.16	2.65
t OTHER DEP	Other Departments																				
School 1				-.15	1.71			-.15	1.93											.24	5.88
School 2		-.30	9.72	.15	1.56	-.24	5.30	.16	2.00					-.42	16.00	-.29	7.74	-.34	10.85		
School 3				-.18	1.18							.12	1.14	-.22	3.89			-.18	2.91	.35	7.05
School 4		-.15	2.31	.09	0.60	-.16	2.64							-.31	9.24			-.25	5.12		
School 5										.14	1.29	.29	7.65	-.11	1.0.					.25	5.29
School 6		.27	5.29			.20	3.92	-.12	1.22	.29	5.60	.37	10.46							.25	5.91
School 7														-.18	3.13			-.21	3.80		
School 8																					
School 9																					
Average	SIRNNQ 25 Class Size is Appropriate															-.12	1.36				
Responses	SIRNNQ 26 Major vs. Minor vs Elective Field									.23	2.10										
for	SIRNNQ 27 Course is Required	-.18	2.69					.18	1.96	-.23	2.67			-.11	1.17	.16	1.65				
Students	SIRNNQ 28 Expected Grade	-.13	1.46																		
Rating	SIRNNQ 29 Grade Point Average			.24	2.72					-.06	0.21					-.19	2.78			-.20	2.16
This	SIRNNQ 30 Class Level (Freshman-Graduate)			.20	2.28	-.14	2.08			-.20	1.92										
TA	SIRNNQ 31 Sex			-.16	1.79			.21	3.47			.16	2.63								

Use of English in student-initiated classroom interchanges (Q42). This question ("When the instructor responded to student questions or statements in class, his or her English-language ability made the answers unclear or difficult to understand") dealt with more spontaneous reception and production than the preceding "lecturing to the class" question (Q40). For the FSI and TSE analyses, the R^2 values were quite high and essentially equivalent (.61 and .60, respectively), with the FSI beta weight somewhat higher (-.57) than that of the TSE (-.45). In both instances, however, the FSI or TSE scores were the best single predictors of student ratings of this aspect of the instructor's English language use.

Use of English in tutorials/laboratory sessions (Q43). This aspect of instructor language use was represented by the question "In individual (one-on-one) teaching situations such as in-class tutorials or laboratory sessions, it was easy for me to understand what the instructor was saying"). Of the language-specific questions, this was the least well predicted in both the FSI and TSE analyses, with identical R^2 values (.46) in both instances. The partial regression coefficients were also virtually equivalent for FSI (B=.48) and TSE (B=.50) scores. Despite the relatively low R^2 for both analyses, FSI and TSE scores were again the highest single predictor of student ratings of instructor performance on this variable.

Use of English in academically-related private conversations (Q44). Student ratings of the instructor on this language-use variable ("When the instructor was talking privately with me about course-related matters [for example, after class or during an office appointment], I had trouble understanding what he or she was saying") were quite predictable in both the FSI (R^2 =.66) and TSE (R^2 =.59) analyses, although for this particular question, the beta weight of the FSI scores (-.54) was noticeably higher than that of the TSE scores (-.37). This result may be associated with the reasonable assumption that one-on-one communication in this less structured and more "conversational" setting would be somewhat more sensitively measured by the FSI technique (which consists of examiner/examinee interchanges in an actual conversational setting) than by the booklet- and tape-mediated format of the TSE.

Listening comprehension as measured by student speaking strategy required (Q45). In addition to Q41, this question addresses the instructor's listening comprehension. In the latter instance, however, the focus of the question is on any adaptations of the student's own speech that the student felt it was necessary to make in order to communicate properly with the instructor ("When I was talking to the instructor, I had to change my own way of speaking [for example, use simpler words or talk more slowly than usual] to make sure that the instructor understood what I was saying"). This variable was highly predictable in both FSI (R^2 =.73) and TSE (R^2 =.61) analyses, with a beta weight of -.53 for FSI scores and a somewhat lower beta weight (-.37) for TSE scores. As with question 44, the somewhat greater power of the FSI score in predicting the necessity of "real-time" adjustments on the students' part in the level and pace of their own conversation to accommodate limitations in instructor listening comprehension

level, could be attributed to the conversationally-based format of the FSI by comparison to the less flexible and more highly "automated" TSE procedure.

Ratings of individual language elements (pronunciation, grammar, and vocabulary) (Q46-48). Questions 46 through 48 asked the students to rate the extent to which the instructor's "pronunciation of English," "English grammar," and "English vocabulary" interfered with understanding, on a 5-point scale ranging from "did not interfere" to "interfered completely." For both the FSI and TSE analyses, student ratings of these individual aspects of the instructor's spoken language performance were highly predictable, with R^2 values of .62 to .72 for FSI, and .48 to .67 for TSE. For all three questions, and for both FSI and TSE analyses, the beta weights of the FSI and TSE scores were appreciably higher than those of any of the other predictor variables, although the FSI beta weights, in each instance, are higher in absolute terms than the corresponding TSE beta weights (-.70 vs. -.57 for pronunication, -.76 vs. -.44 for grammar, and -.80 vs. -.56 for vocabulary). Again, although in all three instances, the predictive strength of the FSI is, not surprisingly, somewhat higher than that of the TSE--as a probable consequence of the capacity of the FSI procedure to adapt the level and content of the test more flexibly to the performance of the particular examinee being tested--the observed TSE values are quite respectably high and, as indicated, represent the best single predictors from among all the variables included in the TSE regression analysis.

Overall ability to communicate in English (Q49). The final language-specific question used in the study asked the students to evaluate the instructor's "overall ability to communicate in English," as reflected by the ease with which they were able to "understand" the instructor's speech. This global evaluation of speaking proficiency was highly predicted in both FSI and TSE analyses, with R^2 values of .72 and .61, respectively. The beta weight for the contribution of FSI scores was -.69, with all the other beta weights at considerably lower values (.29 to -.05). The corresponding beta weight for the TSE scores was -.63, again markedly higher than the beta values for the other variables included in the analysis (.35 to .11).

In summary, the regressions involving both FSI and TSE scores in the prediction of student ratings of language-specific aspects of instruction provided impressive evidence for the construct validity of both instruments as appropriate measures of "on-the-job" speaking performance of non-native English speaking teaching assistants, and even more impressive evidence of their predictive ability. For both the FSI and TSE analyses, the obtained R^2 values and the high beta weights associated with both test instruments suggest quite strongly that preliminary assessment of non-native English speaking applicants as to English speaking ability, by means of the FSI or TSE, can be considered statistically and conceptually appropriate in connection with their selection for teaching assistant positions. In this regard, measured English speaking ability, as reflected in FSI or TSE scores, is observed to add markedly to the accuracy of prediction available from such background variables as prior teaching experience in English,

residence in English-speaking countries, and formal English study in the native country or in the U.S. Indeed, the moderate negative correlation of years of English study in the native country with ratings of language competency in speaking situations is both revealing and somewhat disturbing. To the extent that formal native country training in English is oriented toward the written, rather than the conversational idiom, such training might be expected to raise TOEFL scores to acceptable levels without concomitant impact on spoken English. Students with extensive English study in their native countries might then be admitted to U.S. institutions without additional screening requirements for spoken English, but might be seriously deficient in the spoken language. The consistency of the negative correlations across the language-related questions for English study in the native country would seem to support this interpretation.

Prediction of Other Instructional Factors

As instruments developed explicitly to measure proficiency in spoken English, as distinguished from the considerably more generalized and more complex set of attributes associated with effective teaching performance, it would not be expected that either the FSI or the TSE would predict "overall teaching performance" on the part of non-native English speaking teaching assistants with the same degree of effectiveness as for English-speaking proficiency per se. However, inasmuch as speaking proficiency in English is almost indisputably viewed as one of the major components of the more generalized concept of "teaching effectiveness," it would be a reasonable expectation that speaking proficiency scores, as provided by the FSI or TSE, would exhibit some degree of relationship to the general "teaching effectiveness" variables provided by the SIR, although not at the same high absolute levels of prediction observed for the language-specific variables as summarized in the preceding section. This assumption was generally borne out in the regression analysis results for the non-language-specific individual questions and global factor scores of the SIR.

Table 8 gives the regression weights, betas, and R^2 values for the prediction of specific SIR questions pertaining to quality of lectures, discussions, and laboratories, and to overall ratings of the effectiveness of the instructor and the overall value of the course.

In the FSI regression analyses, the most predictable of the SIR overall quality ratings for various aspects of instruction (Q35-Q39) were Q36 ("overall quality of class discussions"), with an R^2 value of .47; and Q39 ("effectiveness of this instructor" [relative to other instructors in the student's experience]), with an R^2 of .46. FSI beta weights for these two questions were .42 and .44, respectively. Generally similar results were found for these two questions in the TSE regressions (Q36: R^2 =.42, B=.32; Q39: R^2 =.36, B =.31).

Neither the FSI nor TSE scores were significantly associated with Q35 ("overall quality of the lectures"). Since in several cases, the teaching assistant was not in fact the principal lecturer, this lack of association

Table 8

Regression Analysis--SIR Questions

FSI Analysis			Overall Quality of Lectures		Overall Quality of Class Discussions		Overall Quality of Laboratories		Overall Value of Course		Effectiveness of Instructor	
			Q35		Q36		Q37		Q38		Q39	
			R^2 = .36		R^2 = .47		R^2 = .33		R^2 = .34		R^2 = .46	
			Beta	F	Beta	F	Beta	F	Beta	F	Beta	F
FSI Total	FSI Total				.42	11.83	.15	1.22			.44	12.84
US MONTHS	Months in U.S. or English-speaking Country											
US ENG WKS	Weeks of English Study in U.S.		.31	3.55	.17	2.50			.12	1.00	.19	3.02
US ENG NOW	Currently Taking English Course		-.20	1.54			-.16	1.64				
ENG STUDY	Years of English Study in Native Country						-.29	4.21			-.20	2.95
PRIOR ENG	Prior Course Taught Using English				-.13	1.12						
HIGHEST D	Highest Academic Degree Received										.12	1.10
YRS TEACH	Years Teaching in Any Language				.20	2.64	.22	2.68			.11	0.96
ROLE 1	Teaching vs. Discussion or Lab Assessment				-.19	2.50	.11	0.83				
DEPT ENGINEER	Engineering Department										-.18	2.03
DEPT MATH SCI	Mathematics or Science Department								-.35	7.26		
DEPT FOR LANG	Foreign Language Department				.26	4.06					.31	7.18
DEPT OTHER DEP	Other Departments		.39	8.97								
School 1			-.24	3.60			-.27	4.48				
School 2			.38	8.96	.36	10.54			.32	7.19	.33	8.74
School 3							-.32	6.06				
School 4			.31	5.73					.15	1.51		
School 5							-.31	4.85				
School 6			.23	3.40							.26	4.76
School 7												
School 8												
School 9												
Average	SIRNNQ 25	Class Size is Appropriate			.17	1.96	-.17	1.64	.24	3.76		
Responses	SIRNNQ 26	Major vs. Minor vs Elective Field	-.39	7.49								
for	SIRNNQ 27	Course is Required							-.29	5.46		
Students	SIRNNQ 28	Expected Grade			.15	1.42						
Rating	SIRNNQ 29	Grade Point Average					-.23	2.91				
This	SIRNNQ 30	Class Level (Freshman-Graduate)	.28	3.71								
TA	SIRNNQ 31	Sex	.18	1.99			-.32	6.31			.24	3.91

Table 8 (cont.)

TSE Analysis			Overall Quality of Lectures		Overall Quality of Class Discussions		Overall Quality of Laboratories		Overall Value of Course		Effectiveness of Instructor	
			Q35		Q36		Q37		Q38		Q39	
			R^2 = .36		R^2 = .42		R^2 = .34		R^2 = .34		R^2 = .36	
			Beta	F	Beta	F	Beta	F	Beta	F	Beta	F
TSE COMP	TSE Comprehensibility				.32	7.79	.31	5.10			.31	7.18
US MONTHS	Months in U.S. or English-speaking Country											
US ENG WKS	Weeks of English Study in U.S.		.31	3.55			-.18	1.97	.12	1.00		
US ENG NOW	Currently Taking English Course		-.20	1.54							.15	1.78
ENG STUDY	Years of English Study in Native Country						-.17	1.75				
PRIOR ENG	Prior Course Taught Using English											
HIGHEST D	Highest Academic Degree Received										.17	2.05
YRS TEACH	Years Teaching in Any Language						.18	1.82				
ROLE 1	Teaching vs. Discussion or Lab Assistant						.21	2.27				
Dept	ENGINEER	Engineering Department					.16	1.20			-.14	1.12
Dept	MATH SCI	Mathematics or Science Department			.22	1.29			-.35	7.20		
Dept	FOR LANG	Foreign Language Department			.46	9.23	-.13	0.96			.39	10.95
Dept	OTHER DEP	Other Departments	.39	8.97	.29	3.03						
	School 1		-.24	3.60			-.25	3.74				
	School 2		.38	8.96	.35	8.30			.32	7.19	.30	6.68
	School 3				-.12	1.02						
	School 4						-.17	1.71				
	School 5		.31	5.73			.16	1.40	.15	1.51		
	School 6											
	School 7		.23	3.40			.24	2.81			.26	4.68
	School 8											
	School 9											
Average	SIRNNQ 25	Class Size is Appropriate			.25	4.30			.24	3.76		
Responses	SIRNNQ 26	Major vs. Minor vs Elective Field	-.39	7.49								
for	SIRNNQ 27	Course is Required							-.29	5.46		
Students	SIRNNQ 28	Expected Grade			.21	2.66						
Rating	SIRNNQ 29	Grade Point Average										
This	SIRNNQ 30	Class Level (Freshman-Graduate)	.28	3.71								
TA	SIRNNQ 31	Sex	.18	1.99			-.25	3.14			-.21	2.71

may be attributable to this circumstance. Question 38 ("overall value of the course to me") was also not predicted by either FSI or TSE, no doubt as a consequence of the many other considerations involved in such a generalized judgment on the student's part.

Table 9 shows the regression analyses for prediction of the SIR general factor scores. The results with respect to the contribution of the FSI and TSE scores in the total prediction equation are mixed. FSI scores were significantly related to five of the six factor scores (Course Organization and Planning, Faculty-Student Interaction, Communication, Course Difficulty and Workload, and Tests and Exams), but with fairly low beta weights, ranging from -.37 (for Course Difficulty and Workload) to .17 (Tests and Exams). TSE scores were related significantly to four of the six SIR factors (Course Organization and Planning, Communication, Course Difficulty and Workload, and Tests and Exams). Course Difficulty and Workload is most highly predicted by the TSE (B=-.50), with the other beta weights considerably lower (.12 to .16).

The Course Difficulty and Workload loading suggests that the relationship of language to appropriate difficulty is negative, but because this factor score is "folded" (i.e., "too difficult" and "too easy" ratings both yield low scores), the observed relationship is difficult to interpret.

In summary, except for the Course Difficulty and Workload factor, neither the FSI nor TSE is found to contribute markedly to the prediction of SIR factor scores, although this result would not be wholly unanticipated in view of the fact that non-language aspects of the teacher's performance, such as command of subject matter, personal enthusiasm, organization, and other attributes would be expected to weigh heavily in these very generalized judgments, over and above the contribution of language proficiency as such. For SIR questions more directly related to in-class language behavior (e.g., "overall value of class discussions"), the predictive value of the FSI and TSE scores is appreciably higher, as previously noted.

Table 10 gives the correlations among the variables for the 60 cases included in the regressions. Because of differences in rating baselines across schools and departments (particularly Foreign Language departments), these zero-order correlations are not as meaningful as are the previously-reported beta weights. With few observations (60) and many variables, any variable which enters a regression equation with a sign (positive or negative) different from that in the zero-order correlation matrix is called a "suppressor variable" by optimists. Inspection of Table 10 reveals the variables reported in the preceding regression analyses in fact bore the same sign in the basic correlations as they did in the final regression.

For example, Q49, "Instructor's overall ability to communicate in English interferred with understanding" shows strongest correlations with other student reports of instructor language competence, in particular with Q46 (r=.93) Q47 and Q48 (r=.91) and with Q42 (r=.87), and Q40 (r=.85), suggesting that the dependent variable is highly reliable.

Table 9

Regression Analysis--SIR Factor Scores

FSI Analysis			F1 Course Organization and Planning		F2 Faculty-student Interaction		F3 Communication		F4 Course Difficulty and Workload		F5 Textbooks and Reading Assignments		F6 Tests and Exams	
			R^2 = .33		R^2 = .39		R^2 = .50		R^2 = .55		R^2 = .31		R^2 = .34	
			Beta	F	Beta	F	Beta	F	Beta	F	Beta	F	Beta	F
FSI Total	FSI Total		.27	4..	.21	2.45	.25	3.56	-.37	8.27			.17	1.58
US MONTHS	Months in U.S. or English-speaking Country		.23	3.43	.24	4.67			.40	7.31	.15	1.58	.23	3.41
US ENG WKS	Weeks of English Study in U.S.								-.24	2.61				
US ENG NOW	Currently Taking English Course		-.16	1.54	-.14	1.25			-.24	3.70				
ENG STUDY	Years of English Study in Native Country													
PRIOR ENG	Prior Course Taught Using English		.13	1.00							.20	2.66		
HIGHEST D	Highest Academic Degree Received				.21	3.15			-.12	1.09				
YRS TEACH	Years Teaching in Any Language								-.41	13.36			.24	3.05
ROLE 1	Teaching vs. Discussion or Lab Assistant													
Dept ENGINEER	Engineering Department						-.08	0.31						
Dept MATH SCI	Mathematics or Science Department				.32	7.03	.50	13.03					-.24	2.33
Dept FOR LANG	Foreign Language Department				.11	0.74			.23	2.09			.17	1.58
Dept OTHER DEP	Other Departments		-.15	1.33			-.12	1.10	.28	4.81				
School 1			.41	10.31	.37	9.51	.06	0.28	.36	8.19	.25	3.71	.21	2.65
School 2							-.42	8.55	.32	3.44				
School 3			.16	1.43					-.18	1.30				
School 4													.35	6.92
School 5											-.42	9.44		
School 6														
School 7			.21	2.69									.25	3.32
School 8														
School 9														
Average	SIRNNQ 25	Class Size is Appropriate	.17	1.73										
Responses	SIRNNQ 26	Major vs. Minor vs Elective Field	-.22	3.22			.45	9.61			-.22	1.94	-.54	10.42
for	SIRNNQ 27	Course is Required					-.24	2.89	-.19	2.81	-.14	0.63		
Students	SIRNNQ 28	Expected Grade							.39	8.28	.24	2.42		
Rating	SIRNNQ 29	Grade Point Average					.26	3.48	-.45	6.84	-.39	6.05		
This	SIRNNQ 30	Class Level (Freshman-Graduate)			.12	1.10							.35	5.22
TA	SIRNNQ 31	Sex												

Table 9 (cont.)

TSE Analysis			F1		F2		F3		F4		F5		F6	
			Course Organization and Planning		Faculty-student Interaction		Communication		Course Difficulty and Workload		Textbooks and Reading Assignments		Tests and Exams	
			R^2 = .32		R^2 = .36		R^2 = .48		R^2 = .61		R^2 = .31		R^2 = .32	
			Beta	F	Beta	F	Beta	F	Beta	F	Beta	F	Beta	F
TSE COMP	TSE Comprehensibility		.12	0.72			.14	1.27	-.50	17.15			.16	1.40
US MONTHS	Months in U.S. or English-speaking Country													
US ENG WKS	Weeks of English Study in U.S.		.22	2.96	.26	5.48			.53	13.55	.15	1.58	.21	2.61
US ENG NOW	Currently Taking English Course								-.28	4.12				
ENG STUDY	Years of English Study in Native Country								-.26	5.52				
PRIOR ENG	Prior Course Taught Using English													
HIGHEST D	Highest Academic Degree Received										.20	2.66		
YRS TEACH	Years Teaching in Any Language				.17	2.23								
ROLE 1	Teaching vs. Discussion or Lab Assistant								-.36	12.32			.19	1.88
DEPT	ENGINEER	Engineering Department												
	MATH SCI	Mathematics or Science Department					-.17	1.30						
	FOR LANG	Foreign Language Department			.39	11.54	.50	12.71						
	OTHER DEP	Other Departments	.21	1.22	.21	3.15			.30	4.00			.22	2.65
	School 1		-.20	2.07			-.14	1.35	.25	4.58				
	School 2		.42	10.13	.38	10.97	.09	0.65	.40	11.76	.25	3.71	.25	3.71
	School 3						-.40	7.24	.40	6.24				
	School 4		.10	0.35					-.24	2.53				
	School 5												.39	8.31
	School 6										-.42	9.44		
	School 7		.16	1.62									.21	2.51
	School 8													
	School 9													
Average	SIRNNQ 25	Class Size is Appropriate	.22	2.51										
Responses	SIRNNQ 26	Major vs. Minor vs Elective Field	-.30	4.40			.42	8.11			-.22	1.94	-.41	8.55
for	SIRNNQ 27	Course is Required					-.21	2.09	-.16	2.29	-.13	0.63		
Students	SIRNNQ 28	Expected Grade							.35	7.69	.24	2.42		
Rating	SIRNNQ 29	Grade Point Average					.23	2.56	-.48	9.24	-.39	6.05		
This	SIRNNQ 30	Class Level (Freshman-Graduate)	.15	1.16	.14	1.54							.30	4.02
TA	SIRNNQ 31	Sex												

Table 10

Intercorrelations of Regression Analysis Variables

CORRELATION COEFFICIENTS

VARIABLE	FS1	FS2	FS3	FS4	FS5	FS6	Q35	Q36	Q37	Q38	Q39	Q40	Q41	Q42	Q43	Q44	Q45	Q46	Q47	Q48	Q49	Q25	Q26	Q27
FS1	1.00	0.74	0.11	0.30	0.39	0.54	0.78	0.70	0.51	0.61	0.83	-0.55	0.41	-0.61	0.40	-0.52	-0.40	-0.58	-0.67	-0.56	-0.62	0.13	-0.20	-0.15
FS2	0.74	1.00	0.28	0.41	0.31	0.40	0.59	0.79	0.29	0.54	0.77	-0.63	0.43	-0.65	0.43	-0.51	-0.41	-0.56	-0.61	-0.49	-0.56	0.09	-0.19	-0.20
FS3	0.11	0.28	1.00	-0.08	0.06	0.12	0.36	0.45	0.00	0.26	0.34	-0.33	0.35	-0.38	0.21	-0.19	-0.15	-0.31	-0.18	-0.19	-0.23	0.07	0.08	-0.25
FS4	0.30	0.41	-0.08	1.00	0.39	0.29	0.36	0.34	0.23	0.41	0.28	0.08	-0.14	0.01	-0.13	0.04	0.07	0.06	-0.02	0.02	0.02	0.08	0.00	-0.16
FS5	0.39	0.31	0.06	0.39	1.00	0.57	0.53	0.30	0.36	0.67	0.33	-0.09	-0.02	-0.14	0.05	-0.20	0.04	-0.03	-0.08	-0.01	-0.04	0.11	-0.26	-0.35
FS6	0.54	0.40	0.12	0.29	0.57	1.00	0.63	0.51	0.36	0.50	0.52	-0.27	0.18	-0.28	0.15	-0.34	-0.07	-0.30	-0.29	-0.26	-0.26	0.02	-0.28	-0.22
Q35	0.78	0.59	0.36	0.36	0.53	0.63	1.00	0.70	0.35	0.66	0.74	-0.43	0.39	-0.51	0.33	-0.45	-0.23	-0.40	-0.41	-0.36	-0.44	0.07	-0.25	-0.22
Q36	0.70	0.79	0.45	0.34	0.30	0.51	0.70	1.00	0.31	0.64	0.80	-0.64	0.44	-0.68	0.41	-0.49	-0.38	-0.65	-0.65	-0.52	-0.62	0.22	-0.21	-0.26
Q37	0.51	0.29	0.00	0.23	0.36	0.36	0.35	0.31	1.00	0.41	0.47	-0.27	0.16	-0.28	0.15	-0.24	-0.22	-0.27	-0.31	-0.29	-0.30	-0.12	0.09	-0.03
Q38	0.61	0.54	0.26	0.41	0.67	0.50	0.66	0.64	0.41	1.00	0.63	-0.42	0.23	-0.45	0.22	-0.30	-0.17	-0.38	-0.43	-0.28	-0.38	0.22	-0.13	-0.39
Q39	0.83	0.77	0.34	0.28	0.33	0.52	0.74	0.80	0.47	0.63	1.00	-0.68	0.56	-0.72	0.47	-0.62	-0.52	-0.71	-0.71	-0.65	-0.72	0.11	-0.18	-0.20
Q40	-0.55	-0.63	-0.33	0.08	-0.09	-0.27	-0.43	-0.64	-0.27	-0.42	-0.68	1.00	-0.76	0.93	-0.74	0.74	0.74	0.91	0.85	0.79	0.85	-0.00	0.09	0.05
Q41	0.41	0.43	0.35	-0.14	-0.02	0.18	0.39	0.44	0.16	0.23	0.56	-0.76	1.00	-0.76	0.63	-0.62	-0.79	-0.73	-0.67	-0.77	-0.75	-0.18	0.13	0.14
Q42	-0.61	-0.65	-0.38	0.01	-0.14	-0.28	-0.51	-0.68	-0.28	-0.45	-0.72	0.93	-0.76	1.00	-0.73	0.77	0.78	0.90	0.84	0.82	0.87	0.04	0.05	0.08
Q43	0.40	0.43	0.21	-0.13	0.05	0.15	0.33	0.41	0.15	0.22	0.47	-0.74	0.63	-0.73	1.00	-0.67	-0.59	-0.69	-0.62	-0.60	-0.67	0.02	-0.13	-0.06
Q44	-0.52	-0.51	-0.19	0.04	-0.20	-0.34	-0.45	-0.49	-0.24	-0.30	-0.62	0.74	-0.62	0.77	-0.67	1.00	0.77	0.76	0.73	0.77	0.80	0.12	0.14	0.04
Q45	-0.40	-0.41	-0.15	0.07	0.04	-0.07	-0.23	-0.38	-0.22	-0.17	-0.52	0.74	-0.79	0.78	-0.59	0.77	1.00	0.78	0.72	0.88	0.80	0.20	-0.12	-0.13
Q46	-0.58	-0.56	-0.31	0.06	-0.03	-0.30	-0.40	-0.65	-0.27	-0.38	-0.71	0.91	-0.73	0.90	-0.69	0.76	0.78	1.00	0.90	0.87	0.93	0.06	0.09	0.06
Q47	-0.67	-0.61	-0.18	-0.02	-0.08	-0.29	-0.41	-0.65	-0.31	-0.43	-0.71	0.85	-0.67	0.84	-0.62	0.73	0.72	0.90	1.00	0.91	0.91	0.00	0.08	0.09
Q48	-0.56	-0.49	-0.19	0.02	-0.01	-0.26	-0.36	-0.52	-0.29	-0.28	-0.65	0.79	-0.77	0.82	-0.60	0.77	0.88	0.87	0.91	1.00	0.91	0.15	-0.06	-0.01
Q49	-0.62	-0.56	-0.23	0.02	-0.04	-0.26	-0.44	-0.62	-0.29	-0.38	-0.72	0.85	-0.75	0.87	-0.67	0.80	0.80	0.93	0.91	0.91	1.00	0.08	-0.01	-0.00
Q25	0.13	0.09	0.07	0.08	0.11	0.02	0.07	0.22	-0.12	0.22	0.11	-0.00	-0.18	0.04	0.02	0.12	0.20	0.06	0.00	0.15	0.08	1.00	-0.27	-0.23
Q26	-0.20	-0.19	0.08	0.00	-0.26	-0.28	-0.25	-0.21	0.09	-0.13	-0.18	0.09	0.13	0.05	-0.13	0.14	-0.12	0.09	0.08	-0.06	-0.01	-0.27	1.00	0.59
Q27	-0.15	-0.20	-0.25	-0.16	-0.35	-0.22	-0.22	-0.26	-0.03	-0.39	-0.20	0.05	0.14	0.08	-0.06	0.04	-0.13	0.06	0.09	-0.01	-0.00	-0.23	0.59	1.00
Q28	-0.05	0.20	0.27	0.40	0.09	0.09	0.15	0.16	-0.09	0.23	0.05	0.02	-0.04	0.01	-0.04	0.14	0.21	0.07	0.09	0.16	0.13	-0.05	-0.04	-0.30
Q29	0.10	0.17	0.14	0.09	-0.02	0.02	0.18	0.12	-0.13	0.04	0.12	-0.22	0.11	-0.16	0.19	-0.22	-0.02	-0.18	-0.20	-0.12	-0.16	0.06	-0.08	-0.17
Q30	0.04	0.12	0.21	0.11	-0.07	-0.04	-0.02	0.04	0.11	0.09	0.09	-0.11	0.20	-0.16	0.07	-0.02	-0.14	-0.10	-0.13	-0.13	-0.12	-0.14	0.52	0.10
Q31	0.01	-0.01	0.02	0.10	0.21	0.08	0.21	0.10	-0.28	0.08	-0.08	0.07	-0.28	0.03	0.16	0.06	0.28	0.08	0.11	0.17	0.13	0.18	-0.31	-0.28
MATHSCI	-0.09	-0.29	-0.37	-0.03	-0.14	-0.05	-0.10	-0.24	0.02	-0.32	-0.27	0.43	-0.29	0.43	-0.29	0.38	0.36	0.45	0.40	0.36	0.42	0.10	0.06	0.26
ENGINEER	-0.10	0.02	0.14	0.09	-0.10	-0.04	-0.24	-0.09	0.19	0.05	-0.01	-0.18	0.16	-0.13	0.07	-0.06	-0.23	-0.20	-0.21	-0.24	-0.15	-0.20	0.37	0.14
FORLANG	0.13	0.33	0.42	0.03	0.19	0.03	0.22	0.42	-0.09	0.32	0.35	-0.28	0.04	-0.29	0.30	-0.22	-0.01	-0.31	-0.19	-0.02	-0.19	0.33	-0.52	-0.53
OTHERDEP	0.10	0.09	0.03	-0.06	0.12	0.08	0.15	0.05	-0.11	0.11	0.09	-0.18	0.19	-0.21	0.07	-0.26	-0.24	-0.16	-0.18	-0.24	-0.25	-0.21	0.02	-0.03
SCHOOL1	-0.07	0.13	-0.13	0.19	0.04	0.05	-0.12	0.09	-0.21	0.00	-0.06	-0.03	-0.10	0.01	-0.04	-0.08	-0.07	0.01	-0.09	-0.05	-0.02	0.13	-0.10	0.11
SCHOOL2	0.26	0.27	0.20	0.07	0.07	0.10	0.19	0.24	0.14	0.17	0.24	-0.25	0.27	-0.23	0.08	-0.18	-0.23	-0.24	-0.27	-0.27	-0.21	-0.28	0.14	0.04
SCHOOL3	-0.08	-0.03	-0.09	0.06	-0.03	-0.08	-0.06	-0.01	-0.03	-0.01	-0.01	-0.20	0.00	-0.18	0.26	-0.23	-0.15	-0.20	-0.15	-0.14	-0.14	0.12	-0.12	-0.16
SCHOOL4	0.12	0.02	0.08	-0.25	0.02	-0.08	0.04	-0.04	-0.11	0.02	0.02	-0.06	0.14	-0.11	-0.03	-0.09	-0.15	-0.06	-0.08	-0.12	-0.12	-0.07	0.12	-0.05
SCHOOL5	-0.07	-0.21	-0.09	-0.10	0.18	0.15	0.06	-0.16	0.12	-0.02	-0.19	0.22	-0.14	0.26	-0.22	0.22	0.29	0.34	0.30	0.33	0.37	0.18	-0.01	0.09
SCHOOL6	-0.17	-0.1[illegible]	-0.05	-0.06	-0.30	-0.09	-0.19	-0.07	-0.06	-0.17	-0.08	0.17	0.01	0.20	-0.06	0.19	0.13	0.09	0.16	0.14	0.04	-0.21	-0.03	0.18
SCHOOL7	0.08	0.05	0.12	0.20	0.09	-0.00	0.12	0.03	0.10	0.06	0.15	0.21	-0.24	0.09	-0.10	0.23	0.21	0.11	0.12	0.11	0.09	0.16	0.02	-0.17
TSECOMP	0.18	0.14	0.13	-0.30	-0.08	0.18	0.04	0.29	0.17	0.11	0.29	-0.60	0.52	-0.53	0.51	-0.54	-0.58	-0.68	-0.58	-0.62	-0.68	-0.08	-0.06	0.05
FSITOTAL	0.20	0.26	0.27	-0.22	-0.09	0.13	0.15	0.34	0.06	0.16	0.37	-0.65	0.56	-0.60	0.49	-0.64	-0.63	-0.71	-0.65	-0.69	-0.72	-0.12	-0.08	0.00
ENGSTUDY	-0.08	-0.07	-0.02	-0.16	-0.01	-0.00	-0.03	0.02	-0.11	-0.05	-0.02	-0.02	0.04	-0.03	-0.08	-0.14	-0.11	-0.10	-0.11	-0.12	-0.11	-0.15	0.01	0.12
USMONTHS	0.03	-0.02	-0.20	-0.16	-0.04	-0.05	-0.14	-0.10	0.09	-0.05	0.04	-0.34	0.25	-0.20	0.20	-0.21	-0.41	-0.24	-0.19	-0.31	-0.28	-0.14	0.07	0.19
USENGHOW	0.01	0.07	-0.08	0.13	0.07	0.06	-0.07	-0.05	-0.17	0.02	0.05	0.00	0.05	-0.03	-0.03	0.14	0.02	-0.01	0.06	0.06	-0.01	0.11	0.11	0.08
USENGWKS	0.14	0.22	-0.12	0.20	0.08	0.21	0.03	0.14	-0.15	0.07	0.13	-0.16	0.20	-0.15	0.16	-0.07	-0.13	-0.22	-0.17	-0.16	-0.24	0.12	-0.05	0.11
PRIORENG	0.20	0.09	0.05	-0.27	0.16	0.04	0.16	0.08	0.13	0.17	0.26	-0.32	0.29	-0.31	0.17	-0.36	-0.38	-0.27	-0.23	-0.26	-0.30	0.00	-0.20	-0.04
YRSTEACH	0.23	0.21	-0.00	-0.02	0.17	-0.01	0.12	0.20	0.06	0.17	0.18	0.01	-0.18	-0.02	-0.01	0.18	0.22	0.11	-0.04	0.14	0.05	0.32	-0.25	-0.15
ROLE1	-0.00	0.03	0.06	-0.23	0.13	0.22	-0.04	0.12	0.04	0.12	0.01	-0.17	0.08	-0.13	0.24	-0.23	-0.07	-0.17	-0.10	-0.08	-0.11	-0.09	-0.23	-0.17
HIGHESTD	0.12	0.04	-0.03	-0.00	0.18	0.10	0.08	0.02	-0.08	0.13	0.05	-0.02	0.13	-0.01	-0.09	-0.14	-0.09	0.02	-0.11	-0.17	-0.10	-0.09	0.06	-0.12

Table 10 (cont.)

	Q28	Q29	Q30	Q31	MATHSCI	ENGINEER	FORLANG	OTHERDEF	SCHOOL1	SCHOOL2	SCHOOL3	SCHOOL4	SCHOOL5	SCHOOL6	SCHOOL7	TSECOMP	FSITOTAL	ENGSTUDY	USMONTHS	USENGNOW	USENGWES	PRIORENG	YRSTEACH	ROLE1	HIGHESTD
FS1	-0.05	0.10	0.04	0.01	-0.09	-0.10	0.13	0.10	-0.07	0.26	-0.08	0.12	-0.07	-0.17	0.08	0.18	0.20	-0.08	0.03	0.01	0.14	0.20	0.23	-0.00	0.12
FS2	0.20	0.17	0.11	-0.01	-0.29	0.02	0.33	0.09	0.13	0.27	-0.03	0.02	-0.21	-0.11	0.05	0.14	0.26	-0.07	-0.02	0.07	0.22	0.09	0.21	0.03	0.04
FS3	0.[illegible]	0.14	0.21	0.02	-0.37	0.14	0.42	0.03	-0.13	0.20	-0.09	0.08	-0.09	-0.05	0.12	0.13	0.27	-0.02	-0.20	-0.08	-0.12	0.05	-0.00	0.06	-0.03
FS4	0.40	0.09	0.11	0.10	-0.03	0.09	0.03	-0.06	0.19	0.07	0.06	-0.25	-0.10	-0.06	0.20	-0.30	-0.22	-0.16	-0.16	0.13	0.20	-0.27	-0.02	-0.23	-0.00
FS5	0.09	-0.02	-0.07	0.21	-0.14	-0.10	0.19	0.12	0.04	0.07	-0.03	0.02	0.18	-0.30	0.09	-0.08	-0.09	-0.01	-0.04	0.07	0.08	0.16	0.17	0.13	0.18
FS6	0.09	0.02	-0.04	0.08	-0.05	-0.04	0.03	0.08	0.05	0.10	-0.08	-0.08	0.15	-0.09	-0.00	0.18	0.13	-0.00	-0.05	0.06	0.21	0.04	-0.01	0.22	0.10
Q35	0.15	0.18	-0.02	0.21	-0.10	-0.24	0.22	0.15	-0.12	0.19	-0.06	0.04	0.06	-0.19	0.12	0.04	0.15	-0.03	-0.14	-0.07	0.03	0.16	0.12	-0.04	0.08
Q36	0.16	0.12	0.04	0.10	-0.24	-0.09	0.42	0.05	0.09	0.24	-0.01	-0.04	-0.16	-0.07	0.03	0.29	0.34	0.02	-0.10	-0.05	0.14	0.08	0.20	0.12	0.02
Q37	-0.09	-0.13	0.11	-0.28	0.02	0.19	-0.09	-0.11	-0.21	0.16	-0.03	-0.11	0.12	-0.06	0.10	0.17	0.06	-0.11	0.09	-0.17	-0.15	0.13	0.06	0.05	-0.08
Q38	0.23	0.04	0.09	0.08	-0.32	0.05	0.32	0.11	0.00	0.17	-0.01	0.02	-0.02	-0.17	0.06	0.11	0.16	-0.05	-0.06	0.02	0.07	0.17	0.17	0.12	0.13
Q39	0.05	0.12	0.09	-0.08	-0.27	-0.01	0.35	0.09	-0.06	0.24	-0.01	0.02	-0.19	-0.08	0.15	0.29	0.37	-0.02	0.04	0.05	0.13	0.26	0.18	0.01	0.05
Q40	0.02	-0.22	-0.11	0.07	0.43	-0.18	-0.28	-0.18	-0.03	-0.25	-0.20	-0.06	0.22	0.17	0.21	-0.60	-0.65	-0.02	-0.24	0.00	-0.16	-0.32	0.01	-0.17	-0.02
Q41	-0.04	0.1[illegible]	0.20	-0.28	-0.29	0.16	0.04	0.19	-0.11	0.27	0.00	0.14	-0.14	0.01	-0.24	0.52	0.56	0.04	0.25	0.05	0.20	0.29	-0.18	0.08	0.13
Q42	0.01	-0.16	-0.16	0.03	0.43	-0.13	-0.29	-0.21	0.01	-0.23	-0.18	-0.11	0.26	0.20	0.09	-0.54	-0.60	-0.03	-0.20	-0.03	-0.15	-0.31	-0.02	-0.13	-0.01
Q43	-0.04	0.19	0.07	0.16	-0.29	0.07	0.30	0.07	-0.04	0.08	0.26	-0.03	-0.22	-0.06	-0.10	0.51	0.49	-0.08	0.20	-0.03	0.16	0.17	-0.01	0.24	-0.09
Q44	0.14	-0.22	-0.02	0.06	0.38	-0.06	-0.22	-0.26	-0.07	-0.18	-0.23	-0.09	0.22	0.19	0.23	-0.54	-0.64	-0.14	-0.21	0.14	-0.07	-0.36	0.18	-0.23	-0.14
Q45	0.21	-0.07	-0.14	0.28	0.36	-0.23	-0.01	-0.24	-0.07	-0.23	-0.15	-0.15	0.29	0.13	0.21	-0.58	-0.63	-0.11	-0.41	0.02	-0.13	-0.38	0.22	-0.07	-0.09
Q46	0.07	-0.18	-0.10	0.08	0.45	-0.20	-0.31	-0.16	0.01	-0.24	-0.20	-0.06	0.34	0.09	0.11	-0.68	-0.71	-0.10	-0.24	-0.01	-0.22	-0.27	0.11	-0.17	0.02
Q47	0.09	-0.[illegible]0	-0.13	0.11	0.40	-0.21	-0.19	-0.18	-0.09	-0.27	-0.15	-0.08	0.30	0.16	0.12	-0.58	-0.65	-0.11	-0.19	0.06	-0.17	-0.23	-0.04	-0.10	-0.11
Q48	0.16	-0.12	-0.13	0.17	0.36	-0.24	-0.02	-0.24	-0.05	-0.27	-0.14	-0.12	0.33	0.14	0.11	-0.62	-0.69	-0.12	-0.31	0.06	-0.16	-0.26	0.14	-0.08	-0.17
Q49	0.13	-0.16	-0.12	0.13	0.41	-0.15	-0.19	-0.25	-0.02	-0.21	-0.14	-0.12	0.37	0.04	0.09	-0.68	-0.71	-0.11	-0.28	-0.01	-0.24	-0.30	0.05	-0.11	-0.10
Q25	-0.05	0.06	-0.14	0.18	0.10	-0.20	0.33	-0.21	0.13	-0.26	0.12	-0.07	0.18	-0.21	0.16	-0.08	-0.12	-0.15	-0.14	0.11	0.12	0.00	0.32	-0.09	-0.09
Q26	-0.06	-0.08	0.52	-0.31	0.06	0.37	-0.52	0.02	-0.10	0.14	-0.12	0.12	-0.01	-0.03	0.02	-0.06	-0.08	0.01	0.07	0.11	-0.05	-0.20	-0.25	-0.23	0.06
Q27	-0.30	-0.17	0.10	-0.28	0.26	0.14	-0.53	-0.03	0.10	0.04	-0.16	-0.05	0.09	0.18	-0.17	0.05	0.00	0.12	0.19	0.08	0.11	-0.04	-0.15	-0.17	-0.12
Q28	1.00	0.49	0.31	0.15	-0.27	0.23	0.27	-0.06	-0.12	0.02	0.19	-0.16	-0.08	-0.01	0.07	-0.20	-0.13	-0.23	-0.42	0.14	0.06	-0.34	-0.08	-0.05	-0.10
Q29	0.49	1.00	0.20	0.21	-0.24	0.02	0.21	0.12	-0.07	0.06	0.62	-0.05	-0.29	-0.40	-0.08	-0.00	-0.08	-0.18	-0.21	0.02	0.03	-0.26	-0.06	-0.03	0.17
Q30	0.31	0.20	1.00	-0.27	-0.23	0.53	-0.05	-0.11	-0.07	0.17	0.08	0.08	-0.24	-0.06	0.05	-0.03	0.01	-0.11	-0.05	0.09	-0.01	-0.26	-0.22	-0.27	-0.01
Q31	0.15	0.21	-0.27	1.00	0.08	-0.44	0.26	0.06	0.02	-0.17	0.22	-0.09	-0.04	-0.15	0.18	-0.12	-0.08	-0.10	-0.19	0.08	0.05	-0.18	0.14	0.16	-0.02
MATHSCI	-0.27	-0.24	-0.23	0.08	1.00	-0.43	-0.39	-0.59	-0.27	0.10	-0.22	-0.32	0.38	0.20	0.09	-0.43	-0.52	0.03	-0.08	-0.02	-0.09	-0.19	-0.02	-0.08	-0.13
ENGINEER	0.23	0.02	0.53	-0.44	-0.43	1.00	-0.12	-0.18	0.15	0.19	0.09	-0.13	-0.16	-0.02	-0.10	0.19	0.26	0.06	0.12	0.08	0.08	-0.02	-0.29	-0.08	-0.14
FORLANG	0.27	0.21	-0.05	0.26	-0.39	-0.12	1.00	-0.17	-0.08	-0.12	0.25	-0.12	-0.15	0.00	0.13	0.09	0.18	-0.05	-0.23	0.02	0.05	0.15	0.09	0.17	-0.21
OTHERDEF	-0.06	0.12	-0.11	0.06	-0.59	-0.18	-0.17	1.00	0.27	-0.18	0.02	0.60	-0.22	-0.22	-0.13	0.31	0.30	-0.05	0.17	-0.05	0.01	0.13	0.20	0.04	0.43
SCHOOL 1	-0.12	-0.07	-0.07	0.02	-0.27	0.15	-0.08	0.27	1.00	-0.08	-0.16	-0.08	-0.10	-0.10	-0.06	0.18	0.14	0.10	0.23	0.09	0.28	-0.05	0.12	0.11	0.08
SCHOOL 2	0.02	0.06	0.17	-0.17	0.10	0.19	-0.12	-0.18	-0.08	1.00	-0.25	-0.13	-0.16	-0.16	-0.10	0.06	0.10	0.23	-0.15	-0.13	-0.18	0.09	-0.01	0.05	0.06
SCHOOL 3	0.19	0.62	0.08	0.22	-0.22	0.09	0.25	0.02	-0.16	-0.25	1.00	-0.25	-0.30	-0.30	-0.18	0.18	0.15	-0.02	0.13	0.01	0.05	-0.11	-0.18	-0.12	-0.01
SCHOOL 4	-0.16	-0.05	0.08	-0.09	-0.32	-0.13	-0.12	0.60	-0.08	-0.13	-0.25	1.00	-0.16	-0.16	-0.10	-0.05	-0.04	-0.20	0.02	0.08	-0.06	0.19	0.20	0.05	0.38
SCHOOL 5	-0.08	-0.29	-0.24	-0.04	0.38	-0.16	-0.15	-0.22	-0.10	-0.16	-0.30	-0.16	1.00	-0.20	-0.12	-0.28	-0.25	-0.06	-0.23	0.03	-0.09	0.04	-0.01	0.00	-0.07
SCHOOL 6	-0.01	-0.40	-0.06	-0.15	0.20	-0.02	0.00	-0.22	-0.10	-0.16	-0.30	-0.16	-0.20	1.00	-0.12	0.09	0.05	0.02	0.18	-0.06	0.14	-0.04	-0.14	0.11	-0.16
SCHOOL 7	0.07	-0.08	0.05	0.18	0.09	-0.10	0.13	-0.13	-0.06	-0.10	-0.18	-0.10	-0.12	-0.12	1.00	-0.22	-0.18	-0.03	-0.21	0.02	-0.10	-0.11	0.26	-0.37	-0.26
TSECOMP	-0.20	-0.00	-0.03	-0.12	-0.43	0.19	0.09	0.31	0.18	0.06	0.18	-0.05	-0.28	0.09	-0.22	1.00	0.79	0.22	0.37	-0.09	0.24	0.30	-0.08	0.17	-0.00
FSITOTAL	-0.13	-0.00	0.01	-0.08	-0.52	0.26	0.18	0.30	0.14	0.10	0.15	-0.04	-0.25	0.05	-0.18	0.79	1.00	0.35	0.23	-0.25	0.03	0.32	-0.17	0.05	-0.01
ENGSTUDY	-0.23	-0.18	-0.11	-0.10	0.03	0.06	-0.05	-0.05	0.10	0.23	-0.02	-0.20	-0.06	0.02	-0.03	0.22	0.35	1.00	0.06	-0.28	-0.12	0.12	-0.11	-0.17	-0.01
USMONTHS	-0.42	-0.21	-0.05	-0.19	-0.07	0.12	-0.23	0.17	0.23	-0.15	0.13	0.02	-0.23	0.18	-0.21	0.37	0.23	0.06	1.00	0.15	0.31	0.33	-0.06	0.17	0.12
USENGNOW	0.16	0.02	0.09	0.08	-0.02	0.08	0.02	-0.05	0.09	-0.13	0.01	0.08	0.03	-0.06	0.02	-0.09	-0.25	-0.28	0.15	1.00	0.65	-0.04	0.11	0.05	-0.10
USENGWES	0.06	0.03	-0.01	0.05	-0.09	0.07	0.05	0.01	0.28	-0.18	0.05	-0.06	-0.09	0.14	-0.10	0.24	0.03	-0.12	0.31	0.65	1.00	0.00	-0.00	0.22	-0.03
PRIORENG	-0.34	-0.26	-0.26	-0.17	-0.19	-0.02	0.15	0.13	-0.05	0.09	-0.11	0.19	0.04	-0.04	-0.11	0.30	0.32	0.12	0.33	-0.04	0.00	1.00	0.29	0.12	-0.00
YRSTEACH	-0.08	-0.06	-0.22	0.14	-0.02	-0.29	0.09	0.20	0.12	-0.03	-0.18	0.20	-0.03	-0.14	0.26	-0.08	-0.17	-0.11	-0.06	0.11	-0.00	0.29	1.00	-0.06	0.12
ROLE 1	-0.05	-0.03	-0.27	0.16	-0.08	-0.08	0.17	0.04	0.11	0.05	-0.02	0.05	0.00	0.11	-0.37	0.17	0.05	-0.17	0.17	0.05	0.22	0.12	-0.06	1.00	0.07
HIGHESTD	-0.10	0.17	-0.01	-0.02	-0.13	-0.14	-0.21	0.43	0.08	0.06	-0.01	0.38	-0.07	-0.16	-0.26	-0.00	-0.01	-0.01	0.12	-0.10	-0.03	-0.00	0.12	0.07	1.00

Among the potential predictors of Q49 (FSI or TSE; Q25-31; plus department, school, and instructor background variables), FSI total (r=-.71) and TSE comprehensibility (r=-.68) show by far the strongest correlation with this overall "ability to communicate" rating. The negative signs are in the expected direction, with high predictor scores corresponding to low ("did not interfere with understanding") mean class rating values. The only other predictors with correlations of at least .30 in absolute value are Math or Science departments (r=.41), yielding less favorable ratings; school 5 (r=.37), also rating instructors more stringently; and prior teaching experience in English (r=-.30), predicting higher ratings. Other, weaker, predictors include months in the U.S. (r=-.28); "other" department (r=-.25); weeks of English study in the U.S. (r=-.24); school 2 (r=-.21); Foreign Language department (r=-.19); and grade-point average (r=-.16). In the regression analysis reported in Table 7, the strongest predictors of Q49 are FSI total (B=-.69); School 1 (B=.25); grade-point average (B=-.25); and school 5 (B=.24). The only sign reversal is for school 1 (r=-.02, Table 10), and reflects the contrast between this essentially zero correlation and the negative correlations for schools 2, 3, and 4.

In a similar manner, one can compare the weights of each regression equation with the original correlations, and conclude that the method employed (adding variables to the regression until the shrunken R^2 began to decrease) was successful in identifying and correcting for meaningful sources of rating variability. The values of the adjusted squared multiple correlations for each regression are given in Appendix G.

Native-English Speaking Cohort Analysis

The fact that this was an observational study, and that the participating TA's were not (nor could they have been) randomly assigned to schools, departments, or instructional roles makes it necessary to interpret with caution the contribution of background variables to the predictions. Students at School 2, for example, gave significantly more favorable ratings to their instructors on most questions, and this "school" variable entered a large number of the regressions as a significant predictor even when test scores were taken into account. Schools 6 and 7 tended to give lower ratings. Foreign language departments gave higher ratings, and Engineering and Math/Science students lower ratings on many variables.

Instructors serving in the role of teacher, rather than of discussion leader or laboratory assistant, received higher ratings on some variables. These differences may reflect actual differences in language level confounded with the background variables. School 2, for example, may in fact have more effective non-native English-speaking TA's than do other schools, or the differences may reflect different rating standards on the students' part.

In the original design of the study, a matched native-speaking TA was to be rated for each non-native speaking subject, in an effort to estimate and correct for such differences in rating behavior. Because it was not

possible to obtain matching data for one-third of the usable non-native subjects, and because the sample size was already small, the native English-speaking controls were not included in the regression analyses as such. However, the SIR scores of the 51 native English-speaking matched controls were analyzed separately to determine the degree of similarity of the relations of background variables to ratings in the two groups. If School 2 and Foreign Language departments also rated native TA's more favorably, for example, it would suggest that students in these categories may have a general tendency to give favorable ratings. If ratings of native English-speaking TA's show different patterns from those of non-natives, however, it may be that language ability was confounded with background variables in a way that distorted the relationships in this sample. Table 11 shows the correlation of school with ratings on two representative questions: "overall effectiveness of the instructor" (Q39) and "instructor's overall ability to communicate in English interfered with understanding" (Q49).

Table 11

Correlations of School with Q39 and Q49
for Native and Non-native English Speaking TA's

	Q39 Overall Effectiveness	Q49 Ability to Communicate Interfered
Native English	- .13	.28
Non-Native English	- .06	.22

The similarity of these correlations suggests that school effects applied to non-native English speakers in a manner similar to that for native speakers, and probably represented real school differences in rating behavior.

Table 12 gives correlations of department with instructor effectiveness for the two groups.

Table 12

Correlations of Overall Effectiveness of
Instructor (Q39) with Department

	Math/Science	Engineering	Foreign Language	Other
Native English	-.10	.04	.21	-.07
Non-Native English	-.26	-.04	.37	.08

Foreign language students give the highest ratings to both groups, and Math/Science students the lowest, suggesting a systematic relationship between department and students' rating behavior. However, in this case, the relationships are stronger for non-native English speakers to the degree that we may suspect that non-native Foreign Language teachers may in fact be relatively better teachers of their subject matter, beyond the general tendency of students in their department to give favorable ratings.

Table 13 gives correlations of student ratings of the instructor's "overall ability to communicate in English" (Q49) with department for the two groups. Here the rating is based on the degree of interference with understanding, so that a negative correlation represents a more favorable rating in that department.

Table 13

Correlations of Overall Ability to Communicate in English (Q49) with Department

	Math/Science	Engineering	Foreign Language	Other
Native English	.17	-.07	-.10	-.07
Non-Native English	.43	-.16	-.25	-.21

Here the patterns are identical for the two groups, but the relationship is much stronger in the non-native English group. This notably stronger relationship is partly an artifact of the lower variance of native English speakers on language-related questions (most were at the "no-interference" extreme, with a mean of 1.08, where 1 = "did not interfere" and 5 = "interfered completely": the mean for non-native English TA's was 2.12). The pattern of correlations from most favorable (for Foreign Language) to least favorable (for Math/Science) is, however, identical for the two groups, suggesting that the differences are to some extent in the students, and not solely due to confounding of language ability with department in this particular sample of non-native teaching assistants.

Correlations with instructional role present a somewhat less consistent picture across groups. Table 14 gives the relation of role (1 = course teacher, 2 = discussion section leader, 3 = assistant in laboratory sessions, 5 = grade papers and examinations, 6 = assist professor in research, 7 = other) to the overall effectiveness ratings.

Table 14

Correlations of Overall Effectiveness of Instructor (Q39) with Instructional Role

Role:	1	2	3	4	5	6	7
Native English	.23	-.35	-.07	-.03	-.10	.22	-.16
Non-Native English	.00	-.08	.11	.23	-.03	.28	-.04

The correlations for native English speakers suggest that being the principal teacher, rather than a discussion leader, is associated with a considerable bonus in effectiveness rating. If this "halo" tendency carries over to non-native TA's, it is evidently offset by considerably lower effectiveness in the actual teaching role, which is rated only slightly more favorably than that of discussion leader. Non-native TA's who also reported role 4 (tutorial sessions) or role 6 (assist professor in research) were also rated more favorably. This latter relation was replicated in the native English group, and suggests that students continue to be appreciative of feedback.

In the non-native regressions, being an incumbent of role 1, taking test scores into account, was associated with higher ratings, even though its zero-order correlation with this criterion is also zero in magnitude. This suggests that the tendency to rate the primary teacher higher was still present for non-native English TA's, but was offset by actual teaching deficiencies in this group.

Table 15 gives correlations of overall ability to communicate in English (Q49) with role.

Table 15

Correlations of Overall Ability to Communicate in English (Q49) with Instructional Role

Role:	1	2	3	4	5	6	7
Native English	-.38	.00	.18	.06	.04	.05	-.05
Non-Native English	-.13	.03	-.12	-.21	-.04	-.30	.08

Again, native English-speaking incumbents of role 1 receive markedly more favorable (lower interference) ratings, but non-native English

teachers receive only slightly more favorable assessments. Roles 4 and 6 (in addition to some combination of roles 1, 2, or 3) again are associated with more favorable ratings for non-native English teaching assistants.

In summary, the data available from a group of native-English speaking teaching assistants in the same departments and schools of a subset of the target sample suggests that school and department variations cut across language background of the teaching assistant, but that language and teaching ability may have been confounded to some extent with teaching role in the target sample in such a way as to make the relationship of teaching role to ratings less comparable between the native and non-native samples. However, when language test scores are taken into account in the regression equations, the advantage associated with occupying Role 1 that appears for native speakers also becomes apparent in the non-native group.

The contribution to R^2 of school and department is thus more likely to represent a generalizable phenomenon, at least for these departments in these schools, but that of teaching role may be to some extent an artifact of the particular distribution of ability across instructional roles in the present sample.

Sample Expectancy Table

Table 16 gives an example of an expectancy table constructed from the responses of the 28 science and math instructors (the largest departmental category among the subjects with complete data) to Question 49, "The instructor's overall ability to communicate in English interfered with understanding."

A perfect relationship might manifest itself in a pattern similar to that illustrated below, given the TSE score distribution of this group, and that the cut points were appropriate:

Numbers of Teaching Assistants

	2	0	0		4.0 interfered considerably
	0	14	0		
	0		12		1.0 did not interfere
TSE:	90	155	220	285	

Percentage of Total in Each Third of Score Range

100	0	0
0	100	0
0	0	100

In fact, Table 16 shows that the relationship was not perfect for either predictor. It does, however, illustrate that 93 percent of those with TSE scores below 220 were rated as having an ability to communicate

which interfered slightly, somewhat, or considerably with understanding, whereas 58 percent of those with TSE scores above 220 were rated as having an ability to communicate which interfered slightly or less with understanding, while only 8 percent (1 individual) of this latter group was rated as having an overall ability to communicate that interfered "somewhat" with understanding.

Because of the small number of cases, this table is for illustrative purposes only, although the relationship it illustrates does significantly depart from chance expectation.

Table 16

Expectancy Table-Science
and Math Instructors

Instructor's Overall Ability to Communicate
Interfered with Understanding

Numbers of Teaching Assistants

FSI				TSE		
			4.0 Interfered considerably			
			3.4			3.4
5	3	1	3.0 Interfered somewhat	2	6	1
			2.7			2.7
4	4	1		0	7	4
			2.0 Interfered slightly			2.0
0	5	3		0	1	7
			1.3			1.3
1.4	2.4	3.4 4.4	1.0 Did not interfere	90	155	220 285

Percentage of Total in Each
Third of Score Range

			4.0 Interfered considerably			
			3.4			3.4
56	25	20	3.0 Interfered somewhat	100	43	8
			2.7			2.7
44	33	20		0	50	33
			2.0 Interfered slightly			2.0
0	42	60		0	7	58
			1.3			1.3
1.4	2.4	3.4 4.4	1.0 Did not interfere	90	155	220 285

SUMMARY

As outlined at the beginning of this report, the two major purposes of the study were, first, to conduct a concurrent validation analysis of the Test of Spoken English, in its present operational form, using as an external criterion the Foreign Service Institute interview procedure; and second, to obtain use-validation information for one important application of TSE score data--in conjunction with the selection or assignment of non-native English speaking teaching assistants for classroom lecturing or other responsibilities involving active use of spoken English.

For the concurrent validation portion of the study, a total of 134 foreign teaching assistants distributed among nine participating institutions were administered the Test of Spoken English, and of these, 94 were also tested using the FSI interview and associated rating scale.

Interrater reliability coefficients for the four TSE scores (Comprehensibility, Pronunciation, Grammar, and Fluency), based on the correlations of two independent ratings of each test tape, ranged from .77 to .85. These figures may be considered probable underestimates of the scoring reliability of the TSE within the operational testing program, since reported scores are routinely based on the averaged results of two separate ratings of each tape.

The correlational data also provide some evidence of discriminant and convergent validity of the TSE scoring scales, in that each of the four TSE scores was found to correlate somewhat more highly with "itself" than with any of the other scores. Correlations of the averages of the two individual ratings of each TSE tape indicate that the general Comprehensibility rating is more closely related to Pronunciation and Fluency scores than to Grammar, an outcome which is consistent with the intended measurement purpose of the TSE and the test development procedures used in designing the test.

Scoring reliabilities of the FSI were also determined. Although the scoring reliability of the FSI global rating was equivalent to that of the TSE Comprehensibility score (.79), the individual FSI scores for Pronunciation, Grammar, Vocabulary, Fluency, and Comprehension (which are not routinely obtained or reported as part of the regular FSI scoring process) showed consistently lower reliabilities than the Pronunciation, Grammar, and Fluency scores provided by the TSE, especially for Pronunciation, which was considerably more reliably measured by the TSE. In addition, the intercorrelations among the FSI subscores, as based on averaged ratings, showed no consistent patterning that would support the conceptual or operational distinctiveness of the FSI subscores.

A related correlational analysis, which requires some caution in interpretation due to the small number of cases (31), indicates that the TSE Comprehensibility score and the TSE Pronunciation, Grammar, and Fluency scores correlate more highly with the FSI global score than do

either the TOEFL total score or any of the three TOEFL subscores (Listening Comprehension, Vocabulary and Structure, and Reading Comprehension). TSE Grammar is more closely associated with TOEFL total (as well as with the three TOEFL subscores) than are TSE Pronunciation and TSE Fluency. This suggests that TSE Pronunciation and TSE Fluency measure somewhat different aspects of the examinee's language performance than do either the TOEFL (or its subscores) or the Grammar score of the TSE.

On the basis of the results obtained in this study, it would appear that the Test of Spoken English demonstrates very satisfactory levels of scoring reliability for both the overall Comprehensibility score and for the separate diagnostic scores for Pronunciation, Grammar, and Fluency. These latter scores, as provided through the TSE, appear to be more reliable and to exhibit higher discriminant validity than the corresponding subscores of the FSI. Within the interpretative limitations of a rather small subsample of study participants having recent TOEFL scores, administration of the TSE is seen to provide additional reliable information for examinee speaking skills (especially with respect to pronunciation and general fluency) that is not provided by the TOEFL scores themselves.

To carry out the second major (use-validity) analysis for the study, the scoring results of the participating teaching assistants on both the TSE and FSI were used, separately, as predictors of student ratings of the teaching assistants' communicative proficiency in English in classroom lecturing and other instructional situations, as well as of more general aspects of their instructional performance (e.g., use of class time, preparation and organization, general teaching "effectiveness," etc.). Multiple regression techniques were used to relate these criterion variables to the predictor variables of FSI and TSE scores as well as other predictors derived from personal background data (e.g., amount of prior English language study, prior teaching experience, length of time in the United States); and, for control purposes, nominal variables representing institution and department affiliation of the teaching assistant, as well as such instructor-independent variables as appropriateness of class size, whether or not the course was required, grade-point average of the students in the particular course, and others.

Based on a relatively modest sample size (N=60), both FSI and TSE scores were found to be strong predictors of the teaching assistants' communicative performance in English in classroom lecture settings and in in-class question-answer situations, as well as of their communicative effectiveness in one-on-one conversational situations such as student-teacher interchanges in tutorial or laboratory sessions or in after-class or office-visit settings.

In general, FSI scores were found to be slightly more highly predictive of communicative effectiveness than the TSE scores, although in both instances, the absolute magnitudes of the beta weights for these two tests were consistently appreciably higher than the beta weights associated with the other predictors used in the analysis, including such biographical data as length of residence in the United States or other English-speaking

countries, amount of English study in the United States, and prior study of English in the teaching assistant's home country. It may be quite strongly inferred from these analyses that administration of either the FSI or TSE to applicants for teaching assistant positions can provide appreciably greater prediction of their probable communicative performance in English-speaking situations associated with their instructional assignments than is available through biographical data concerning the nature and amount of their prior English study.

With respect to the prediction of more general aspects of teaching performance (e.g., "overall effectiveness" of the instructor or of broad Student Instructional Report factors such as "course organization and planning"), the predictive power of both the FSI and TSE is appreciably reduced by comparison to that for questions addressed specifically to spoken language use in academic settings, a finding that is in keeping with the probable substantial contribution of a large number of additional personality, subject-matter knowledge, and other factors that would be expected to influence "teaching effectiveness" in addition to English speaking proficiency per se. Nonetheless, the consistent pattern of at least moderately high beta values for both TSE and FSI in predicting more generalized aspects of teaching performance is in keeping with the presumed partial contribution of the instructors' English language proficiency to these global performance ratings.

To provide some indication of the possible effects on the regression analysis results of inter-institution or inter-department variations in rating behavior on the part of the students in evaluating their instructors, the SIR ratings given to 51 native English speaking "cohort" teaching assistants were compared to those given to the non-native English teaching assistants in the same institutions and departments. Correlations of institution and department codings with student ratings of "overall effectiveness of the instructor" and with "overall ability to communicate in English" were calculated and compared for the native English and non-native English instructor groups. These results indicate that school effects in the rating of both "overall effectiveness of the instructor" and "ability to communicate in English" operated in a generally similar manner for both the native English and non-native English groups, but that at the department level the observed inter-departmental differences in ratings given to the native English and non-native English instructors may be attributable, at least to some extent, to differences in the students' rating behavior that are beyond the interpretative capability of the present study. It should be emphasized, however, that interaction effects of this type may be considered of relatively minor significance in the interpretation of the overall regression analysis results, in view of the observed high degree of predictive power, as reflected in the zero-order correlations, of both the TSE and FSI as indicators of "communicative proficiency" in English-medium instructional situations.

REFERENCES

Centra, John A. The Student Instructional Report: Its Development and Uses (SIR Report No. 1). Princeton, N.J.: Educational Testing Service, 1972.

Centra, John A. "Student Ratings of Instruction and Their Relationship to Student Learning." In Centra, John A., Two Studies on the Validity of the Student Instructional Report. Princeton, N.J.: Educational Testing Service, 1976, pp. 1-15.

Centra, John A. & Creech, F. Reid. "The Relationship Between Student, Teacher, and Course Characteristics and Student Ratings of Teacher Effectiveness." In Centra, John A., Two Studies on the Validity of the Student Instructional Report. Princeton, N.J.: Educational Testing Service, 1976, pp. 17-74.

Clark, John L. D. & Swinton, Spencer S. An Exploration of Speaking Proficiency Measures in the TOEFL Context (TOEFL Research Report No. 4). Princeton, N.J.: Educational Testing Service, 1979.

Educational Testing Service. Student Instructional Report. Princeton, N.J.: Educational Testing Service, 1971.

Educational Testing Service. Test of Spoken English. Princeton, N.J.: Educational Testing Service, 1980.

Sollenberger, Howard E. "Development and Current Use of the FSI Oral Interview Test." In John L. D. Clark, ed., Direct Testing of Speaking Proficiency: Theory and Application. Princeton, N.J.: Educational Testing Service, 1978, pp. 3-12.

Wilds, Claudia P. "The Oral Interview Test." In Jones, Randall L. & Spolsky, Bernard, eds., Testing Language Proficiency. Arlington, Va.: Center for Applied Linguistics, 1975, pp. 29-44.

APPENDIX A

FSI Level Descriptions

Level	Verbal Descriptions
1	Able to satisfy routine travel needs and minimum courtesy requirements. Can ask and answer questions on topics very familiar to him; within the scope of his very limited language experience can understand simple questions and statements, allowing for slowed speech, repetition or paraphase; speaking vocabulary inadequate to express anything but the most elementary needs; errors in pronunciation and grammar are frequent, but can be understood by a native speaker used to dealing with foreigners attempting to speak his language; while topics which are "very familiar" and elementary needs vary considerably from individual to individual, any person at the S-1 level should be able to order a simple meal, ask for shelter or lodging, ask and give simple directions, make purchases, and tell time.
2	Able to satisfy routine social demands and limited work requirements. Can handle with confidence but not with facility most social situations including introductions and casual conversations about current events, as well as work, family, autobiographical information; can handle limited work requirements, needing help in handling any complications or difficulties; can get the gist of most conversations on non-technical subjects (i.e. topics which require no specialized knowledge) and has a speaking vocabulary sufficient to express himself simply with some circumlocutions; accent, though often quite faulty, is intelligible; can usually handle elementary constructions quite accurately but does not have thorough or confident control of the grammar.
3	Able to speak the language with sufficient structural accuracy and vocabulary to participate effectively in most formal and informal conversations on practical, social, and professional topics. Can discuss particular interests and special fields of competence with reasonable ease; comprehension is quite complete for a normal rate of speech; vocabulary is broad enough that he rarely has to grope for a word; accent may be obviously foreign; control of grammar good; errors never interfere with understanding and rarely disturb the native speaker.
4	Able to use the language fluently and accurately on all levels normally pertinent to professioanl needs. Can understand and participate in any conversation within the range of his experience with a high degree of fluency and precision of vocabulary; would rarely be taken for a native speaker, but can respond appropriately even in unfamiliar situations; errors of pronunciation and grammar quite rare; can handle informal interpreting from and into the language.

5 Speaking proficiency equivalent to that of an educated native speaker. Has complete fluency in the language such that his speech on all levels is fully accepted by educated native speakers in all of its features, including breadth of vocabulary and idiom, colloquialisms, and pertinent cultural references.

FSI "Grid" Rating Form

Factors in Speaking Proficiency

	S-1	S-2	S-3	S-4	S-5
Pronunciation	Often unintelligible	Usually foreign but rarely unintelligible	Sometimes Foreign but always intelligible	Sometimes foreign but always intelligible	Native
Grammar	Accuracy limited to set expressions; almost no control of syntax; often conveys wrong information	Fair control of most basic syntactic patterns; conveys meaning accurately in simple sentences most of time	Good control of most basic syntactic patterns; always conveys meaning accurately in reasonably complex sentences	Makes only occasional errors, and these show no pattern of deficiency	Control equal to that of an educated native speaker
Vocabulary	Adequate only for survival, travel, and basic courtesy needs	Adequate for simple social conversation and routine job needs	Adequate for participation in all general conversation and for professional discussions in a special field	Professional and general vocabulary broad and precise, appropriate to occasion	Equal to vocabulary of an educated native speaker
Fluency	Except for memorized expressions, every utterance required enormous obvious effort	Usually hesitant; often forced to silence by limitations of grammar and vocabulary	Rarely hesitant; always able to sustain conversation through circumlocutions	Speech on all professional matters as apparently effortless as in English; always easy to listen to	Speech at least as fluent as in English on all occasions
Comprehension	May require much repetition, slow rate of speech; understands only very simple, short familiar utterances	In general understands non-technical speech directed to him, but sometimes misinterprets or needs utterances reworded. Usually cannot follow conversation between native speakers	Understands most of what is said to him; can follow speeches, clear radio broadcasts, and most conversation between native speakers, but not in great detail	Can understand all educated speech in any moderately clear context; occasionally baffled by colloquialisms and regionalisms	Equal to that of the native speaker

Appendix B

STUDENT INSTRUCTIONAL REPORT

SIR Report Number

This questionnaire gives you an opportunity to express anonymously your views of this course and the way it has been taught. Indicate the response closest to your view by blackening the appropriate oval. Use a soft lead pencil (preferably No. 2) for all responses to the questionnaire. Do not use an ink or ball point pen.

SECTION I Items 1-20. Blacken one response number for each question.

NA (0) = Not Applicable or don't know. The statement does not apply to this course or instructor, or you simply are not able to give a knowledgeable response.
SA (4) = Strongly Agree. You strongly agree with the statement as it applies to this course or instructor.
A (3) = Agree. You agree more than you disagree with the statement as it applies to this course or instructor.
D (2) = Disagree. You disagree more than you agree with the statement as it applies to this course or instructor.
SD (1) = Strongly Disagree. You strongly disagree with the statement as it applies to this course or instructor.

		NA	SA	A	D	SD
1.	The instructor's objectives for the course have been made clear	⓪	④	③	②	①
2.	There was considerable agreement between the announced objectives of the course and what was actually taught	⓪	④	③	②	①
3.	The instructor used class time well	⓪	④	③	②	①
4.	The instructor was readily available for consultation with students	⓪	④	③	②	①
5.	The instructor seemed to know when students didn't understand the material	⓪	④	③	②	①
6.	Lectures were too repetitive of what was in the textbook(s)	⓪	④	③	②	①
7.	The instructor encouraged students to think for themselves	⓪	④	③	②	①
8.	The instructor seemed genuinely concerned with students' progress and was actively helpful	⓪	④	③	②	①
9.	The instructor made helpful comments on papers or exams	⓪	④	③	②	①
10.	The instructor raised challenging questions or problems for discussion	⓪	④	③	②	①
11.	In this class I felt free to ask questions or express my opinions	⓪	④	③	②	①
12.	The instructor was well-prepared for each class	⓪	④	③	②	①
13.	The instructor told students how they would be evaluated in the course	⓪	④	③	②	①
14.	The instructor summarized or emphasized major points in lectures or discussions	⓪	④	③	②	①
15.	My interest in the subject area has been stimulated by this course	⓪	④	③	②	①
16.	The scope of the course has been too limited; not enough material has been covered	⓪	④	③	②	①
17.	Examinations reflected the important aspects of the course	⓪	④	③	②	①
18.	I have been putting a good deal of effort into this course	⓪	④	③	②	①
19.	The instructor was open to other viewpoints	⓪	④	③	②	①
20.	In my opinion, the instructor has accomplished (is accomplishing) his or her objectives for the course	⓪	④	③	②	①

SECTION II Items 21-31. Blacken one response number for each question.

21. For my preparation and ability, the level of difficulty of this course was:
- ① Very elementary
- ② Somewhat elementary
- ③ About right
- ④ Somewhat difficult
- ⑤ Very difficult

22. The work load for this course in relation to other courses of equal credit was:
- ① Much lighter
- ② Lighter
- ③ About the same
- ④ Heavier
- ⑤ Much heavier

23. For me, the pace at which the instructor covered the material during the term was:
- ① Very slow
- ② Somewhat slow
- ③ Just about right
- ④ Somewhat fast
- ⑤ Very fast

24. To what extent did the instructor use examples or illustrations to help clarify the material?
- ① Frequently
- ② Occasionally
- ③ Seldom
- ④ Never

Questionnaire continued on the other side.

Copyright © 1971 by Educational Testing Service. All Rights Reserved.
No part of the Student Instructional Report may be adapted or reproduced in any form without permission in writing from the publisher.

572MRC39P200X

283562

MRC FORM NO. 6884

25. Was class size satisfactory for the method of conducting the class?

① Yes, most of the time
② No, class was too large
③ No, class was too small
④ It didn't make any difference one way or the other

26. Which one of the following best describes this course for you?

① Major requirement or elective within major field
② Minor requirement or required elective outside major field
③ College requirement but not part of my major or minor field
④ Elective not required in any way
⑤ Other

27. Which one of the following was your most important reason for selecting this course?

① Friend(s) recommended it
② Faculty advisor's recommendation
③ Teacher's excellent reputation
④ Thought I could make a good grade
⑤ Could use pass/no credit option
⑥ It was required
⑦ Subject was of interest
⑧ Other

28. What grade do you expect to receive in this course?

① A
② B
③ C
④ D
⑤ Fail
⑥ Pass
⑦ No credit
⑧ Other

29. What is your approximate cumulative grade-point average?

① 3.50-4.00
② 3.00-3.49
③ 2.50-2.99
④ 2.00-2.49
⑤ 1.50-1.99
⑥ 1.00-1.49
⑦ Less than 1.00
⑧ None yet

30. What is your class level?

① Freshman
② Sophomore
③ Junior
④ Senior
⑤ Graduate
⑥ Other

31. Sex:

① Female
② Male

SECTION III Items 32-39. Blacken one response number for each question.

		Not applicable, don't know, or there were none.	Excellent	Good	Satisfactory	Fair	Poor
32.	Overall, I would rate the textbook(s)	⑥	⑤	④	③	②	①
33.	Overall, I would rate the supplementary readings	⑥	⑤	④	③	②	①
34.	Overall, I would rate the quality of the exams	⑥	⑤	④	③	②	①
35.	I would rate the general quality of the lectures	⑥	⑤	④	③	②	①
36.	I would rate the overall value of class discussions	⑥	⑤	④	③	②	①
37.	Overall, I would rate the laboratories	⑥	⑤	④	③	②	①
38.	I would rate the overall value of this course to me as	⑥	⑤	④	③	②	①

39. Compared to other instructors you have had (secondary school and college), how effective has the instructor been in this course? (Blacken one response number.)

One of the most effective (among the top 10%)	More effective than most (among the top 30%)	About average	Not as effective as most (in the lowest 30%)	One of the least effective (in the lowest 10%)
⑤	④	③	②	①

SECTION IV Items 40-49. If the instructor provided supplementary questions and response options, use this section for responding. Blacken only one response number for each question.

	NA									
40.	⑥	①	②	③	④	⑤				
41.	⑥	①	②	③	④	⑤				
42.	⑥	①	②	③	④	⑤				
43.	⑥	①	②	③	④	⑤				
44.	⑥	①	②	③	④	⑤				
45.	⑥	①	②	③	④	⑤				
46.	⑥	①	②	③	④	⑤				
47.	⑥	①	②	③	④	⑤				
48.	⑥	①	②	③	④	⑤				
49.	⑥	①	②	③	④	⑤				

If you would like to make additional comments about the course or instruction, use a separate sheet of paper. You might elaborate on the particular aspects you liked most as well as those you liked least. Also, how can the course or the way it was taught be improved? PLEASE GIVE THESE COMMENTS TO THE INSTRUCTOR. 50. Your Native Language

If you have any comments, suggestions, or complaints about this questionnaire (for example, the content or responses available), please send them to: Student Instructional Report, Educational Testing Service, Princeton, New Jersey 08541.

Appendix C

STUDENT INSTRUCTIONAL REPORT

Supplementary Questions

SECTION IV Items 40-50.

General Directions. In this section, we would like you to try to separate your instructor's English-language ability from other important aspects of teaching--such as knowledge of subject matter, difficulty of content, and overall course organization--and to answer the following items in terms of the instructor's English-language ability ONLY.

For questions 40-46, please blacken one response number for each item, according to how often this particular situation occurred. Use the following code and please read each item very carefully before responding.

NA (0) = Not applicable or don't know.

(1) = Rarely or never.

(2) = Occasionally.

(3) = About half the time.

(4) = Frequently.

(5) = Always or almost always.

40. When the instructor was lecturing to the class, his or her English interfered with my understanding of what was being said.

41. The instructor appeared to easily understand questions asked or statements made in class by the students.

42. When the instructor responded to student questions or statements in class, his or her English-language ability made the answers unclear or difficult to understand.

43. In individual (one-on-one) teaching situations such as in-class tutorials or laboratory sessions, it was easy for me to understand what the instructor was saying.

44. When the instructor was talking privately with me about course-related matters (for example, after class or during an office appointment), I had trouble understanding what he or she was saying.

45. When I was talking to the instructor, I had to change my own way of speaking (for example, use simpler words or talk more slowly than usual) to make sure that the instructor understood what I was saying.

(PLEASE TURN THE PAGE.) ——————→

For questions 46-49, please blacken one response number for each item, using the following code:

NA (0) = Not applicable or don't know.

(1) = Did not interfere with understanding.

(2) = Interfered slightly with understanding.

(3) = Interfered somewhat with understanding.

(4) = Interfered considerably with understanding.

(5) = Interfered completely with understanding.

46. The instructor's pronunciation of English....

47. The instructor's English grammar....

48. The instructor's English vocabulary....

49. The instructor's overall ability to communicate in English....

For question 50, please follow the instructions below.

50. In the box next to the number 50 on your answer sheet, please write in the name of your own native language (English, Spanish, Chinese, etc.).

Thank you for answering this questionnaire.

Appendix D

Questionnaire for Participants in the TSE Validation Study

The questions below are for research purposes only, and your individual answers will not be made available to anyone at your institution.

1. Your Name (please print clearly)_____
   Last First

2. Your Native Language (mother tongue)_____

3. In the space provided, please write the number of years you have studied English in school in your native country (for example: 3 years; 7 years). Begin with your very earliest study.

   _____years

4. In the space provided, please write the total number of months you have been in the United States or other English-speaking countries. (For example, if you have been in England for one month and in the United States for two years and three months, you would write down 28 months.)

   _____months

5. Have you taken (or are you taking) any English language course(s) in the United States? (Check one.)

   ( ) No.

   ( ) Yes. If you answered "yes," please give the total number of weeks of English language study you have had in the United States:

   _____weeks

6. Before this semester, have you ever taught any course in which the language of instruction was English (in other words, you had to speak in English in order to do the teaching)? Do not count this semester.

   ( ) No, before this semester, I have not taught any course in which I had to use English.

   ( ) Yes, before this semester, I taught a course in (give name of subject)

   _____in which I had to use English.

7. Not counting this academic year, how many years have you been teaching any subject in any country? (Check one.)

   ( ) I have not taught before. ( ) 1 year ( ) 2 years ( ) 3 or more years

8. What is the name of your academic department at the institution?_____

(PLEASE TURN THE PAGE.)

9. What are your responsibilities this semester? (Please check all that apply.)

( ) I am teaching a course in (give subject)_____. The official title of the course is _____.

( ) I lead a discussion section after the professor lectures.

( ) I assist in laboratory sessions (help the students with equipment, answer questions, and so forth).

( ) I discuss their work with individual students (tutorial sessions).

( ) I grade student papers and/or examinations for a professor or another instructor.

( ) I assist a professor in doing research.

( ) Other responsibility (please describe)_____

_____

10. As part of the study, it is necessary for us to use your score on the TOEFL for statistical analysis purposes only. We would therefore request your signature below to indicate your permission for project staff to obtain and use your TOEFL score record only in connection with this study, with the understanding that these scores will be used only by project staff and will not be released to any other persons within or outside of your institution.

_____
Signature

11. On what date did you most recently take the TOEFL?_____
Month Year

12. Where was this TOEFL administered?_____
City/State Country

13. What is the highest academic degree you have received? (Check one.)

( ) B.S., B.A. or equivalent ( ) M.S., M.A. or equiv. ( ) PhD. or equiv.

14. What is your present mailing address (for sending your speaking test scores and other project information)?

_____

_____

_____
City State ZIP

Appendix E

Instructions to Students

(Please read these instructions aloud to the class before handing out the Student Instructional Report sheets and the Supplementary Questions sheets.)

"As part of a study of the relationship between language and instructional effectiveness, we are asking the students in certain classes to fill out a short rating form called the Student Instructional Report, which I will now hand out, along with a sheet containing certain supplementary questions. After handing out these materials, I will explain them in more detail."

(Please distribute to each student a copy of the Student Instructional Report (a white sheet printed in orange) and a copy of the Supplementary Questions (an orange sheet with black printing). When all students have these materials, say:)

"Please look at the top right-hand corner of the Student Instructional Report form. In the space marked "SIR Report Number," please use a lead pencil to write in the number [read aloud the 4-digit number from the left-hand side of the white INSTRUCTOR AND CLASS label on the manila envelope]. If you do not have a pencil, please raise your hand and I will give you one. Leave the left-most box in the SIR Report Number space blank."

(If a blackboard is available, write the number on the board. Then say:)

"Please check that you have accurately written the number [read number again] in the SIR Report Number space.

"Your answers to the questions on the Student Instructional Report are completely anonymous, and you should not put your name on the report form. Because of the anonymous nature of the report, your own answers will not be made available to your instructor or to other persons at the institution (indeed, there is no way to determine who has filled out each answer sheet) and the answers will have no effect whatsoever on your course grade or any other aspects of your course work.

Please be as accurate and as frank as possible in filling out the report form, and give your own personal judgment in each instance. The answers will not be used to evaluate your instructor, and information identifying your instructor will not be released. Therefore, a frank report will benefit the overall teaching at your institution but can neither benefit nor harm individual instructors.

"For the last section of the report (Section IV), you will need to refer to the separate orange sheet entitled "Supplementary Questions," which gives the questions for this section. All of your answers, however, should be marked on the SIR report form.

(OVER)

Please use only a lead pencil in filling out the report form.

If you make a mistake or wish to change an answer, erase your first answer completely, and do not make any stray marks on the report form.

"Answer all questions in terms of your instructor's teaching, lab sessions, or other instructial contacts up to this point in the course. The instructor on which your answers should be based is [give name of instructor from label on manila envelope] and the course is [give title of course].

Please try to answer every question by filling in the appropriate oval on the answer sheet. If a particular question does not apply or if you cannot give a knowledgeable response, mark the 'Not Applicable or Don't Know' oval. Please do not leave any questions blank.

"The entire questionnaire should take approximately 10 to 15 minutes to complete. Thank you for your help, and please let me know if you have any questions."

(After the students have completed the questionnaire, please collect all questionnaires, making sure that the SIR Report Number has been written at the top of the form in each case. The orange "Supplementary Questions" sheets may be discarded, but please put all completed Student Instructional Reports, as well as any extra blank SIR report forms, into the manila envelope and firmly seal the envelope. Please use the white paper seal included in the envelope to supplement the sealing of the manila flap.

It is extremely important that the SIR forms for a given instructor be returned in the specific envelope designated for that instructor, so we would appreciate your checking this matter carefully. The sealed envelope should be returned to the contact person at your institution as soon as possible following administration.)

Thank you for your help!

Appendix F

Student Instructional Report

STUDENT INSTRUCTIONAL REPORT

ETS COLLEGE AND UNIVERSITY PROGRAMS
PRINCETON, NEW JERSEY 08541

EC 0882

All response summaries are a PERCENTAGE of the total number responding, which is: 17

000	1979	81902	4006	1

COURSE ORGANIZATION AND PLANNING	Omits	Not Applicable	4 Strongly Agree	3 Agree	2 Disagree	1 Strongly Disagree	Mean*	Decile Equivalent***
1. The instructor's objectives for the course have been made clear.	0	18	35	41	6	0	3.36	80
2. There was considerable agreement between the announced course objectives and what was actually taught.	0	18	53	24	0	6	3.88	88
3. The instructor used class time well.	0	0	35	41	24	0	3.12	38
12. The instructor was well-prepared for each class.	0	0	35	29	29	6	2.94	28
13. The instructor told students how they would be evaluated in the course.	0	12	6	24	53	0	2.41	18
14. The instructor summarized or emphasized major points in lectures or discussions.	6		35	41	18	0	3.27	48
20. In my opinion, the instructor has accomplished (is accomplishing) his or her objectives for the course.	6	18	29	47	0	0	3.38	60

FACULTY-STUDENT INTERACTION	Omits	Not Applicable	4 Strongly Agree	3 Agree	2 Disagree	1 Strongly Disagree	Mean*	Decile Equivalent***
4. The instructor was readily available for consultation with students.	0	18	24	47	12	0	3.14	28
5. The instructor seemed to know when students didn't understand the material.	0	12	29	41	12	6	3.07	50
6. The instructor seemed genuinely concerned with students' progress and was actively helpful.	0	0	24	59	18	0	3.06	30
9. The instructor made helpful comments on papers or exams.	0	6	29	18	35	12	2.69	30
11. In this class I felt free to ask questions or express my opinions.	0	0	12	59	18	12	2.71	18
19. The instructor was open to other viewpoints.	0	18	12	65	6	0	3.07	30

COMMUNICATION	Omits	Not Applicable	4 Strongly Agree	3 Agree	2 Disagree	1 Strongly Disagree	Mean*	Decile Equivalent***
8. Lectures were too repetitive of what was in the textbook(s).	0	12	0	18	53	18	2.88	50 **
7. The instructor encouraged students to think for themselves.	0	6	24	59	12	0	3.13	30
10. The instructor raised challenging questions or problems for discussion.	0	0	0	65	35	0	2.65	10

24. To what extent did the instructor use examples or illustrations to help clarify the material?	Omits	4 Frequently	3 Occasionally	2 Seldom	1 Never	Mean*	Decile Equivalent***
	6	59	29	6	0	3.56	50

26. I would rate the general quality of the lectures:	Omits	Not Applicable	5 Excellent	4 Good	3 Satisfactory	2 Poor	1 Poor	Mean*	Decile Equivalent***
	12	0	18	53	12	6	0	3.93	50

COURSE DIFFICULTY AND WORKLOAD	Omits	1 Very Elementary	2 Somewhat Elementary	3 About Right	4 Somewhat Difficult	5 Very Difficult
21. For my preparation and ability, the level of difficulty of this course was:	0	0	0	24	53	24

22. The workload for this course in relation to other courses of equal credit was:	Omits	1 Much Lighter	2 Lighter	3 About the Same	4 Heavier	5 Much Heavier
	0	0	6	24	35	35

23. For me, the pace at which the instructor covered the material during the term was:	Omits	1 Very Slow	2 Somewhat Slow	3 Just About Right	4 Somewhat Fast	5 Very Fast
	0	0	12	29	53	6

No means are calculated. The middle response generally is the preferred one for items 21, 22, and 23.

TEXTBOOKS AND READINGS	Omits	Not Applicable	5 Excellent	4 Good	3 Satisfactory	2 Poor	1 Poor	Mean*	Decile Equivalent***
32. Overall, I would rate the textbook(s):	6	6	12	18	35	24	0	3.29	29
33. Overall, I would rate the supplementary readings:	12	47	0	18	6	18	0	3.00	20

TESTS AND EXAMS	Omits	Not Applicable	4 Strongly Agree	3 Agree	2 Disagree	1 Strongly Disagree	Mean*	Decile Equivalent***
17. Examinations reflected the important aspects of the course.	0	6	18	65	6	6	3.00	40

34. Overall, I would rate the quality of the exams:	Omits	No exams, don't know, need data	5 Excellent	4 Good	3 Satisfactory	2 Poor	1 Poor	Mean*	Decile Equivalent***
	6	0	0	35	35	18	6	3.06	20

OVERALL EVALUATIONS	Omits	Not Applicable	5 Excellent	4 Good	3 Satisfactory	2 Poor	1 Poor	Mean*	Decile Equivalent***
25. I would rate the overall value of class discussions:	18	6	12	12	29	24	0	3.18	20
27. Overall, I would rate the laboratories:	18	76	0	0	0	0	6	----	----
28. I would rate the overall value of this course to me as:	12	0	18	35	18	12	6	3.53	30

39. Compared to other instructors you have had (secondary school and college), how effective has the instructor been in this course?	Omits	5 Among the Top 10%	4 Above the Top 25%	3 About Average	2 Among the Bottom 25%	1 Among the Bottom 10%	Mean*	Decile Equivalent***
	12	12	24	41	12	0	3.40	30

ITEMS NOT IN ANY CATEGORIES	Omits	Not Applicable	4 Strongly Agree	3 Agree	2 Disagree	1 Strongly Disagree	Mean*	Decile Equivalent***
15. My interest in the subject area has been stimulated by this course.	0	6	12	53	24	6	2.75	30
16. The scope of the course has been too limited; not enough material has been covered.	0	0	0	12	53	35	1.76	40 **
18. I have been putting a good deal of effort into this course.	0	0	47	29	18	6	3.18	80

* OMITS and NOT APPLICABLE responses are excluded in computing mean.
** See reverse side for discussion of items 8 and 16.
*** Deciles based on appropriate two-year or four-year college comparative data. See reverse side.

Copyright © 1979 by Educational Testing Service. All rights reserved.

Comparative Data Tables

The comparative data in the tables on this page were compiled from SIR administrations at two-year colleges and technical institutions and at four-year colleges and universities in the United States and Canada. All item means are distributed at decile intervals and are displayed in numerical order, not grouped by factors. The center column contains the 50th percentile or *median* — that is, for each item half the class means are higher and half are lower than the one in the center column. Similarly, in the 70th percentile column, 30 percent of the class means for each item are higher and 70 percent are lower, whereas, in the 30th percentile column, 70 percent of the class means for each item are higher and 30 percent are lower.

Comparative data are updated every two years by type of college.

Comparative Data for Two-Year Colleges and Technical Institutions
Distribution of the Means by Percentiles
(Based on responses from 196,170 students in 5,960 classes, 1975-1978)

ITEM	90%	80%	70%	60%	Median 50%	40%	30%	20%	10%
1.	3.73	3.64	3.57	3.50	3.43	3.36	3.29	3.19	3.03
2.	3.67	3.55	3.47	3.40	3.33	3.27	3.19	3.09	2.94
3.	3.77	3.67	3.59	3.50	3.44	3.35	3.25	3.13	2.93
4.	3.75	3.65	3.58	3.50	3.44	3.38	3.30	3.20	3.05
5.	3.55	3.43	3.33	3.26	3.19	3.10	3.00	2.88	2.69
6.**	2.55	2.35	2.23	2.13	2.05	2.00	1.90	1.82	1.68
7.	3.67	3.56	3.48	3.41	3.34	3.27	3.20	3.10	3.00
8.	3.72	3.61	3.52	3.44	3.36	3.28	3.19	3.07	2.88
9.	3.60	3.45	3.33	3.25	3.16	3.07	3.00	2.83	2.63
10.	3.62	3.50	3.39	3.30	3.22	3.13	3.04	2.93	2.74
11.	3.75	3.67	3.58	3.50	3.44	3.37	3.29	3.19	3.00
12.	3.80	3.71	3.63	3.57	3.50	3.43	3.33	3.24	3.08
13.	3.69	3.59	3.52	3.45	3.40	3.33	3.25	3.15	3.00
14.	3.71	3.61	3.54	3.47	3.40	3.33	3.25	3.15	3.00
15.	3.64	3.50	3.39	3.29	3.19	3.09	3.00	2.82	2.60
16.**	2.30	2.13	2.03	1.95	1.89	1.82	1.75	1.67	1.57
17.	3.50	3.39	3.31	3.23	3.14	3.07	3.00	2.87	2.67
18.	3.54	3.42	3.33	3.25	3.19	3.11	3.05	2.96	2.82
19.	3.64	3.54	3.45	3.38	3.31	3.24	3.17	3.06	2.88
20.	3.75	3.64	3.57	3.50	3.44	3.36	3.27	3.17	3.00
For items 21-23, the third response is the preferred one.									
24.	3.89	3.81	3.73	3.67	3.60	3.52	3.42	3.31	3.12
For items 25-31, means are not appropriate and are not computed.									
32.	4.33	4.14	4.00	3.89	3.78	3.65	3.50	3.31	3.00
33.	4.25	4.00	3.91	3.79	3.67	3.56	3.42	3.25	3.00
34.	4.27	4.11	4.00	3.90	3.80	3.68	3.55	3.40	3.13
35.	4.60	4.44	4.32	4.20	4.08	3.94	3.78	3.57	3.25
36.	4.47	4.29	4.14	4.00	3.87	3.72	3.57	3.36	3.04
37.	4.50	4.29	4.09	4.00	3.80	3.63	3.50	3.20	2.83
38.	4.57	4.40	4.27	4.16	4.04	3.91	3.78	3.60	3.33
39.	4.54	4.36	4.22	4.08	3.93	3.80	3.65	3.44	3.15

**For items 6 and 16, a higher mean and percentile are usually less desirable, and a lower mean and percentile are generally more desirable or "better."

Comparative Data for Four-Year Colleges and Universities
Distribution of the Means by Percentiles
(Based on responses from 188,000 students in 8,180 classes, 1976-1978)

ITEM	90%	80%	70%	60%	Median 50%	40%	30%	20%	10%
1.	3.71	3.60	3.50	3.43	3.35	3.27	3.18	3.06	2.88
2.	3.64	3.53	3.44	3.36	3.29	3.21	3.12	3.00	2.88
3.	3.73	3.63	3.53	3.43	3.33	3.22	3.09	2.92	2.68
4.	3.72	3.61	3.53	3.45	3.38	3.30	3.20	3.09	2.92
5.	3.50	3.34	3.25	3.16	3.07	2.97	2.86	2.71	2.50
6.**	2.50	2.29	2.16	2.06	2.00	1.90	1.83	1.73	1.60
7.	3.67	3.55	3.44	3.36	3.29	3.20	3.12	3.00	2.85
8.	3.69	3.56	3.45	3.35	3.26	3.15	3.05	2.91	2.70
9.	3.55	3.36	3.22	3.10	3.00	2.86	2.73	2.57	2.35
10.	3.60	3.46	3.35	3.25	3.15	3.06	2.95	2.83	2.64
11.	3.71	3.60	3.51	3.43	3.36	3.28	3.19	3.07	2.90
12.	3.81	3.71	3.62	3.55	3.47	3.38	3.28	3.14	2.91
13.	3.68	3.57	3.50	3.42	3.34	3.27	3.18	3.07	2.89
14.	3.67	3.56	3.47	3.40	3.31	3.23	3.14	3.00	2.83
15.	3.58	3.42	3.29	3.18	3.07	2.95	2.81	2.64	2.40
16.**	2.26	2.09	2.00	1.91	1.84	1.77	1.70	1.62	1.50
17.	3.50	3.36	3.27	3.18	3.09	3.00	2.91	2.77	2.55
18.	3.58	3.44	3.33	3.25	3.17	3.08	3.00	2.90	2.75
19.	3.58	3.46	3.36	3.29	3.21	3.13	3.06	2.93	2.77
20.	3.70	3.59	3.50	3.42	3.33	3.25	3.15	3.00	2.82
For items 21-23, the third response is the preferred one.									
24.	3.88	3.79	3.71	3.64	3.57	3.48	3.38	3.26	3.07
For items 25-31, means are not appropriate and are not computed.									
32.	4.29	4.06	3.92	3.78	3.64	3.50	3.38	3.11	2.78
33.	4.20	4.00	3.83	3.70	3.57	3.43	3.26	3.00	2.73
34.	4.17	4.00	3.83	3.71	3.58	3.44	3.29	3.07	2.75
35.	4.54	4.35	4.19	4.04	3.89	3.72	3.53	3.25	2.88
36.	4.35	4.13	3.97	3.81	3.66	3.48	3.29	3.06	2.71
37.	4.50	4.25	4.00	3.83	3.63	3.45	3.20	3.00	2.50
38.	4.47	4.27	4.12	4.00	3.85	3.70	3.53	3.31	3.00
39.	4.50	4.28	4.11	3.94	3.79	3.62	3.42	3.18	2.83

**For items 6 and 16, a higher mean and percentile are usually less desirable, and a lower mean and percentile are generally more desirable or "better."

Additional Comparative Data

Much more detailed comparative information is available in the *SIR Comparative Data Guide*, a copy of which was sent to your institution with these SIR reports. Data are presented in the *Guide* both in standard SIR report format for ease of comparison and by percentile distribution of the means. Separate *Guides* have been prepared for two-year colleges and four-year colleges. Each *Guide* contains data analyzed for:

- type of institution (two-year or four-year)
- size of class
- level of class (freshman/sophomore and junior/senior — in the four-year *Guide* only)
- type of class (lecture, discussion, lab)
- subject areas — using the subject areas listed on the Instructor's Cover Sheet, data are available for approximately 30 different academic disciplines (prepared separately for two-year and four-year institutions)

See the publications list below for information about ordering additional copies of the *Guide*.

STUDENT INSTRUCTIONAL REPORT

ETS COLLEGE AND UNIVERSITY PROGRAMS
PRINCETON, NEW JERSEY 08541

EC 0802

All response summaries are a PERCENTAGE of the total number responding, which is: 17

CLASS ENROLLMENT	ADMIN. DATE	SIR REPORT NO.	BATCH NUMBER
000	1079	01802	4086

2

STUDENT AND COURSE DESCRIPTIVE ITEMS

	Omits	Yes, most of the time.	No, class was too large.	No, class was too small.	It didn't make any difference.
25. Was class size satisfactory for the method of conducting the class?	6	71	24	0	0

26. Which **one** of the following best describes this course for you?	Omits	Major requirement or elective within major field	Minor requirement or required elective outside major field	College requirement but not part of any major or minor field	Elective not required in any way	Other
	6	82	12	0	0	0

27. Which **one** of the following was your most important reason for selecting this course?	Omits	Friends recommended it.	Faculty advisers recommended it.	Teacher's excellent reputation	Thought I could make a good grade.	Could use passfail credit option.	It was required.	Subject was of interest.	Other
	12	6	18	0	0	0	41	18	6

28. What grade do you expect to receive in this course?	Omits	A	B	C	D	Fail	Pass	No Credit	Other
	12	35	29	12	6	0	0	0	6

29. What is your approximate cumulative grade-point average?	Omits	3.50 - 4.00	3.00 - 3.49	2.50 - 2.99	2.00 - 2.49	1.50 - 1.99	1.00 - 1.49	Less than 1.00	None yet - freshman or transfer
	12	24	6	0	0	0	0	0	59

30. What is your class level?	Omits	Freshman	Sophomore	Junior	Senior	Graduate	Other
	6	88	6	0	0	0	0

LOCAL OPTION ITEMS

	Omits	Not Applicable	1	2	3	4	5	6	7	8	9
40.	12	0	82	0	0	0	6	0	0	0	0
41.	12	0	6	6	0	12	65	0	0	0	0
42.	12	0	76	6	0	0	6	0	0	0	0
43.	12	29	6	0	0	6	47	0	0	0	0
44.	12	24	53	6	0	6	0	0	0	0	0

	Omits	Not Applicable	1	2	3	4	5	6	7	8	9
45.	24	6	65	6	0	0	0	0	0	0	0
46.	24	6	71	0	0	0	0	0	0	0	0
47.	24	6	71	0	0	0	0	0	0	0	0
48.	24	6	71	0	0	0	0	0	0	0	0
49.	29	6	59	6	0	0	0	0	0	0	0

COURSE ORGANIZATION AND PLANNING (Includes Items: 1, 2, 3, 12, 13, 14, 20)

This factor describes the extent to which a teacher makes course objectives clear to students, teaches toward these objectives, and, according to students, is accomplishing the objectives. Higher scores also suggest a well-organized teacher who prepares for each class, summarizes the major points in lectures or discussions, and tells students how they will be evaluated.

0 10% 20% 30% x40% 50% 60% 70% 80% 90% 100%
FACTOR SCORE: 9.72
(SCORE RANGE: 3.73 TO 12.22)

FACULTY/STUDENT INTERACTION (Includes Items: 4, 5, 7, 8, 9, 10, 11, 19) †

This dimension reflects an instructor who is concerned with student progress in the course and seems aware of when students need help. Students feel free to ask questions and to consult with the teacher. Teachers with higher scores are seen as open-minded, challenging, and helpful to students.

0 10% x20% 30% 40% 50% 60% 70% 80% 90% 100%
FACTOR SCORE: 9.07
(SCORE RANGE: 4.44 TO 12.39)

COMMUNICATION (Includes Items: 6, 7, 10, 24, 35) †

Good communication means, according to the items in this dimension, that the teacher's lectures are not overly repetitive of textbook material. Instructors raise challenging questions or problems for discussion, and they use examples or illustrations to help clarify course materials. These characteristics encourage students to think for themselves and, in general, result in lectures of high quality, according to students.

0 10% 20% 30% 40% x50% 60% 70% 80% 90% 100%
FACTOR SCORE: 9.86
(SCORE RANGE: 4.39 TO 13.03)

COURSE DIFFICULTY AND WORKLOAD (Includes Items: 21, 22, 23)

Higher scores on this factor indicate that the difficulty level, workload, and pace of the course are viewed generally by students as about right. Teachers who receive lower scores should go back to the original items to determine whether students see the course as too easy or too difficult or the pace as very slow vs. very fast.

0 x10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
FACTOR SCORE: 8.22
(SCORE RANGE: 4.17 TO 12.49)

TEXTBOOKS AND READINGS (Includes Items: 32,33)

This factor summarizes the extent to which students give favorable ratings to the textbook and supplementary readings. Teachers with lower ratings may want to interview students or to include additional questions of their own the next time they administer SIR to determine which supplementary readings are rated poorly or which aspects of the textbook the students did not like.

0 10% 20% x30% 40% 50% 60% 70% 80% 90% 100%
FACTOR SCORE: 6.27
(SCORE RANGE: 2.93 TO 13.42)

TESTS AND EXAMS (Includes Items: 17, 34)

This factor represents the extent to which students rate course examinations favorably and think that the exams deal with important aspects of the course. Teachers with lower scores will need to determine which other features of the examinations need improving; possibilities include questions that are too vague, exams that are too long or too difficult, and grading that is inconsistent or unrealistic.

0 10% 20% 30% x40% 50% 60% 70% 80% 90% 100%
FACTOR SCORE: 9.64
(SCORE RANGE: 4.26 TO 13.00)

† Items 7 and 10 apply to both Faculty/Student Interaction and to Communication. See the discussion of factor scores on the reverse side of this report.

Copyright © 1979 by Educational Testing Service. All rights reserved.

Percentile Equivalents on the SIR Report

The percentile equivalents appearing *on the front of this report* are in the right-hand column. They have been rounded up or down to the nearest decile. The percentile data used on each report are appropriate for the type of institution (two-year college/technical institution or four-year college/university) in which the instructor *for whom this report was prepared* is teaching. That is, if the instructor is teaching at a two-year college or technical institution, the percentile equivalents printed on that instructor's report will be from two-year institutional comparative data.

Percentile Distribution of SIR Means

The tables on the back of page 1 of this report give instructors information to aid in interpreting their SIR reports. Student ratings typically tend to be favorable. For example, on the 5-point SIR scale (Excellent = 5 to Poor = 1), a mean of 3.6 is numerically above average, but, in comparison with other SIR means, it may be average or even slightly below. It is important to have comparative data to help interpret a report fully. Displaying means as percentile equivalents has proved to be a useful aid in that interpretation.

The comparative data in these tables, and on the report itself, are based on national use of SIR. Equally important and useful are comparative data based on use at the individual institution. Colleges may have such local comparative data prepared through the SIR Combined Report Service.

Concerning the Number of Students Responding

A report for a class with either a small number of students or a small proportion of the class responding should be interpreted with caution. In general, it is desirable to have:

- more than 10 students responding
- at least two thirds of the class completing the forms, unless a smaller proportion is based on a random sample of the students.

The degree of accuracy for each item mean increases as the number of students responding increases. For example, for 10 students, the estimated reliability for the item dealing with the rating of teacher effectiveness (#39) is .78; for 20 students, it is .88; for 25 students, it is .90. See *SIR Report No. 3* for a further discussion of reliability.

To alert you to these reliability concerns, you may find one or more of the following:

- Your report is flagged "See back of page 2: The Number Responding" if (1) 10 or fewer students responded or (2) less than 60 percent of the class responded. (This calculation is based on the information provided on the Instructor's Cover Sheet about class enrollment.)
- If 50 percent or more of the students did not respond to an item or marked it "not applicable," no mean or percentile equivalent is reported.
- If fewer than five students responded, that is, if fewer than five completed answer sheets were received for a class, the responses are not tabulated.

Factor Scores

Factor analysis summarizes student responses to SIR by grouping items of similar content and providing scores for each group of items, that is, for each factor. Since items within each of the six factors tend to be related, a teacher will be rated generally the same on the items that contribute to a factor. For example, if an instructor's score on a factor is above average, the ratings on most of the items in that factor should be above average. Occasionally, items will be in more than one factor, such as items 7 and 10 of SIR, which appear in two factors.

Teachers who receive a low score on a factor should look closely at the responses to the individual items in that factor. At the next SIR administration they could consider adding other items that might examine in more detail that dimension of their teaching. Section IV (supplementary items 40-49) can be used for this purpose. Page 4 of the *Instructor's Guide for Using the SIR* provides a list of suggested items. These items, or others written locally, also can be used to get student reactions to aspects of instruction or the course not included in SIR.

PUBLICATIONS

A number of publications dealing with the broad subject of evaluation and improvement of teaching are available. Some are concerned specifically with the Student Instructional Report and may be helpful in understanding and interpreting your report—for example, *SIR Report No. 4* and *SIR Comparative Data Guide*.

Some are more general and include extensive bibliographies *(Strategies for Improving College Teaching* and *SIR Report No. 1*).

Others are essentially technical, dealing with methodological questions *(Between, Within, and Total Group Factor Analyses of Student Ratings of Instruction* and *Student Points of View in Ratings of College Instruction)*.

Any of the publications in the list at the right may be ordered from the address at the bottom of this page. Please include payment with your order.

SIR Report 1 - The Student Instructional Report: Its Development and Uses ($2)

SIR Report 2 - Two Studies on the Utility of Student Ratings for Improving Teaching ($3)

1. The Effectiveness of Student Feedback in Modifying College Instruction (Also in: *The Journal of Educational Psychology* 65 (1973): 395-401; and [in a condensed version] *Change Magazine*, Volume 5/Number 3/April 1973).
2. Self-Ratings of College Teachers: A Comparison with Student Ratings (Also in: *The Journal of Educational Measurement* 10 [1973]: 287-295.)

SIR Report 3 - The Student Instructional Report ($3)

1. Comparisons with Alumni Ratings
2. Item Reliabilities
3. The Factor Structure

SIR Report 4 - Two Studies on the Validity of the Student Instructional Report ($4)

1. Student Ratings of Instruction and Their Relationship to Student Learning
2. The Relationship between Student, Teacher, and Course Characteristics and Student Ratings of Teacher Effectiveness.

SIR Comparative Data Guide ($5). Described fully on the back of page 1 of this report.

Please indicate whether you wish the *SIR Comparative Data Guide* for two-year or four-year colleges—both are available.

Between, Within, and Total Group Factor Analyses of Student Ratings of Instruction by Robert L. Linn, University of Illinois, John A. Centra, ETS, and Ledyard R. Tucker, University of Illinois. ETS Research Bulletin 74-39. ($2) Also in: *Multivariate Behavioral Research*, July 1975.

Colleagues as Raters of Classroom Instruction (also compares student and colleague ratings on selected SIR items). ETS Research Bulletin 74-18. ($2) Also in: *The Journal of Higher Education*, May/June 1975.

Faculty Development Practices in U.S. Colleges and Universities by John A. Centra, ETS Project Report 76-30. ($2)

The Influence of Different Directions on Student Ratings of Instruction by John A. Centra. ($2)

Strategies for Improving College Teaching, (1972) ERIC Report No. 8. No longer available from AAHE; available as a reprint from ETS. ($2)

Student Points of View in Ratings of College Instruction by John A. Centra, ETS, and Robert L. Linn, University of Illinois. ETS Research Bulletin 73-60. ($2)

The Student as Godfather? The Impact of Student Ratings on Academia. ETS Research Memorandum 73-8. ($2) Also in: *Educational Researcher*, Oct. 1973.

Student Instructional Report
ETS College and University Programs
Box 2813
Princeton, New Jersey 08541

Appendix C

Adjusted Squared Multiple Correlations for Each Regression in Tables 7-9

Variable	Adjusted R^2 TSE	FSI
F S 1	.176	.198
F S 2	.288	.299
F S 3	.387	.413
F S 4	.495	.404
F S 5	.215	.215
F S 6	.209	.204
Q 35	.226	.226
Q 36	.327	.371
Q 37	.176	.179
Q 38	.260	.260
Q 39	.277	.354
Q 40	.528	.603
Q 41	.418	.529
Q 42	.455	.534
Q 43	.367	.365
Q 44	.468	.575
Q 45	.535	.611
Q 46	.598	.661
Q 47	.426	.540
Q 48	.487	.632
Q 49	.543	.639

Following are tables which summarize a more conservative approach to the regression predictions. Because of the small number of observations, the inclusion of all significant predictors runs the risk of overfitting. As a check all regressions were examined at the stage at which the first five most important predictors had entered the regression. Examination of the signs of these predictors in the following tables shows that the sign patterns were similar for TSE and FSI and that these variables retained the signs with which they entered, as other variables were added in the more complete regressions reported in the text.

Appendix G (Continued)

Regression Summaries:
Signs First Five Variables Entering FSI and TSE Regression

	Q40		Q41		Q42		Q43		Q44		Q45		Q46		Q47		Q48		Q49	
	FSI R=.77	TSE R=.73	FSI R=.70	TSE R=.64	FSI R=.72	TSE R=.69	FSI R=.63	TSE R=.64	FSI R=.75	TSE R=.70	FSI R=.76	TSE R=.74	FSI R=.81	TSE R=.78	FSI R=.74	TSE R=.68	FSI R=.78	TSE R=.71	FSI R=.79	TSE R=.74
FSI/TSE	+	+	+	+	+	+	+	+	+	+	-	-	-	-	+	+	-	-	-	-
US MONTHS	-	-									-						-	-		
US ENG WKS				+									-		-				-	
US ENG NOW			+																	
ENG STUDY	+				+		-	-							+					
PRIOR ENG							+		-	-		-								
HIGHEST D			+																	
YRS TEACH				-						+		+								
ROLE 1							+		-											
DEPT ENGINEER																				
DEPT MATH SCI															+					
DEPT FOR LANG	-	-			-	-		+					-	-						
DEPT OTHER DEP																				
School 1			-																	+
School 2	-	-		+	-	-		+			-		-	-	-	-	-	-	-	
School 3							+				-			-				-		+
School 4				+		-							-	-			-	-	-	
School 5												+							+	+
School 6		+			+	+			+	+		+			+					+
School 7																				
SIRNNQ 25																-				
SIRNNQ 26																				
SIRNNQ 27																				
SIRNNQ 28																				
SIRNNQ 29									-	-						-	-			
SIRNNQ 31			-					+												

NOTE: SIRNNQ 30 is deleted from this and following tables because this variable was not among the first five to enter any regression.

Appendix G (Continued)

	Q35		Q36		Q37		Q38		Q39	
	FSI R=.47	TSE R=.47	FSI R=.63	TSE R=.61	FSI R=.43	TSE R=.46	FSI R=.57	TSE R=.57	FSI R=.59	TSE R=.57
FSI/TSE			+	+	+	+			+	+
US MONTHS										
US ENG WKS										
US ENG NOW										
ENG STUDY			-							
PRIOR ENG									-	
HIGHEST D										+
YRS TEACH			+		+				+	
ROLE 1					+					
D ENGINEER										
• MATH SCI							-	-		
D FOR LANG			+	+					+	+
t OTHER DEP	+	+		+						
School 1	-	-			-	-				
School 2	+	+	+	+			+	+	+	+
School 3										
School 4					-					
School 5						+	+	+		
School 6										
School 7	+	+				+				+
SIRNNQ 25				+			+	+		
SIRNNQ 26	-	-								
SIRNNQ 27							-	-		
SIRNNQ 28										
SIRNNQ 29										
SIRNNQ 31					-	-				

Appendix G (Continued)

	FS1		FS2		FS3		FS4		FS5		FS6	
	FSI R=.48	TSE R=.48	FSI R=.48	TSE R=.58	FSI R=.64	TSE R=.64	FSI R=.58	TSE R=.61	FSI R=.51	TSE R=.50	FSI R=.47	TSE R=.47
FSI/TSE							-					
US MONTHS												
US ENG WKS	+	+	+	+			+	+	+	+	+	+
US ENG NOW												
ENG STUDY									+	+		
PRIOR ENG												
HIGHEST D			+	+								
YRS TEACH							-	-				
ROLE 1												
D ENGINEER					-	-						
• MATH SCI												
p FOR LANG			+	+	+	+						
t OTHER DEP			+	+							+	+
School 1							+	+				
School 2	+	+	+	+	+	+					+	+
School 3					-	-						
School 4	+	+										
School 5											+	+
School 6									-	-		
School 7	+	+										
SIRNNQ 25												
SIRNNQ 26	-	-			+	+						
SIRNNQ 27									-	-	-	-
SIRNNQ 28							+	+				
SIRNNQ 29									-	-		
SIRNNQ 31												

TOEFL
Research Reports

The Performance of Native Speakers of English on the Test of English as a Foreign Language: Clark, John L.D. Report 1. November 1977.

Discusses the results of the administration of TOEFL to native speakers of English just prior to their graduation from a college-preparatory high school program. Total test score distributions were highly negatively skewed, reinforcing findings of earlier studies that TOEFL is not psychometrically appropriate for discriminating among native speakers of English with respect to English language competence.

An Evaluation of Alternative Item Formats for Testing English as a Foreign Language: Pike, Lewis W. Report 2. June 1979.

Describes an extensive research study conducted from 1972 to 1974 that was designed to explore possible changes in the format and content of TOEFL. Questions of validation, criterion selection, and content specifications were investigated. The report includes the results of these findings and discusses the implications for TOEFL content specifications and internal structure. This study contributed to the restructuring of TOEFL beginning in 1976.

The Performance of Non-Native Speakers of English on TOEFL and Verbal Aptitude Tests: Angelis, Paul J.; Swinton, Spencer S.; and Cowell, William R. Report 3. October 1979.

Gives the results of a study in which 400 graduate and undergraduate applicants took TOEFL, the GRE Verbal or the SAT Verbal, and the Test of Standard Written English (TSWE). Included in the report are comparative data on performance across tests and interpretive information on how combined test results might best be used in the admission process.

An Exploration of Speaking Proficiency Measures in the TOEFL Context: Clark, John L.D., and Swinton, Spencer S. Report 4. October 1979.

Describes a three-year study involving the development and experimental administration of test formats and item types aimed at measuring the English-speaking proficiency of nonnative speakers. Factor analysis and other techniques were used to identify subsets of item formats and individual items having satisfactory correlations with the Foreign Service Institute criterion interview administered to the test subjects. The results were grouped into a prototype "Test of Spoken English."

The Relationship between Scores on the Graduate Management Admission Test and the Test of English as a Foreign Language: Powers, Donald E. Report 5. December 1980.

Summarizes analyses indicating performance of 6,000 nonnative speakers of English on TOEFL and GMAT. In addition to comparisons between native and nonnative speakers, data are included showing performance by language background. A variety of analyses support the basic differences in the two tests by showing expected GMAT verbal scores for various levels of TOEFL scores.

Factor Analysis of the Test of English as a Foreign Language for Several Language Groups: Powers, Donald E., and Swinton, Spencer S. Report 6. December 1980.

Provides evidence from a set of exploratory analytical techniques that three major factors underlie performance on TOEFL. Some support is also found for concluding that these factors may be interpreted differently for several language groups. The report discusses implications for making inferences based on TOEFL subscores and considerations for future test development.

The Test of Spoken English as a Measure of Communicative Ability in English-Medium Instructional Settings: Clark, John L.D., and Swinton, Spencer S. Report 7. December 1980.

Presents the results of a study that examined the performance of foreign teaching assistants on the Test of Spoken English in relation to their classroom performance as judged by students. Also includes, for purposes of comparison, data showing performance of the same groups of teaching assistants on the Foreign Service oral interview and on TOEFL. Based on the analyses conducted in the study, TSE is shown to be a valid predictor of language abilities for nonnative English-speaking graduate teaching assistants.

Effects of Item Disclosure on TOEFL Performance: Angelis, Paul J.; Hale, Gordon A.; and Thibodeau, Lawrence A. Report 8. December 1980.

Reports the findings of a study designed to examine the effects of performance on TOEFL when a subset of items have been disclosed prior to an administration. Based on data from 16 intensive English training programs, the results indicate significant increases in performance in proportion to the number of items made available to students. Details are provided showing separate results by language group and by item type.

The above reports are currently available. Other research reports are planned. For further information about any of the TOEFL Research Reports, write to:

TOEFL Program Office
Box 899
Princeton, NJ 08541, USA